

---

# **Analysen der Schwingungsspektren von Biomolekülen mit Hybridmethoden**

**Galina Babitzki**

---



München 2009



---

# **Analysen der Schwingungsspektren von Biomolekülen mit Hybridmethoden**

**Galina Babitzki**

---

Dissertation  
an der Fakultät für Physik  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Galina Babitzki  
aus Charkow, Ukraine

München, im Juli 2009

Erstgutachter: Prof. Paul Tavan

Zweitgutachter: Prof. Regina de Vivie-Riedle

Tag der mündlichen Prüfung: 21.09.2009



# Zusammenfassung

Hybridmethoden, welche eine quantenmechanische Beschreibung eines Moleküls im Rahmen der Dichtefunktionaltheorie (DFT) mit einer molekülmechanischen (MM) Modellierung seiner Umgebung kombinieren, eröffnen einen Zugang zur Berechnung von Lösungsmittelleffekten in molekularen Schwingungsspektren. In der vorliegenden Arbeit wird eine solche DFT/MM-Hybridmethode exemplarisch zur Analyse der Schwingungsspektren von Biomolekülen eingesetzt. Durch Vergleich mit spektroskopischen Daten werden die erzielbare Genauigkeit und der dazu nötige Rechenaufwand ausgelotet. Bei den untersuchten Systemen handelt es sich um den Retinalchromophor der lichtgetriebenen Protonenpumpe Bakteriorhodopsin (BR) und um ein in Methanol gelöstes, lichtschaltbares *beta*-Hairpinpeptid.

Die inhomogen verbreiterten Schwingungsspektren werden mit DFT/MM unter Verwendung der "instantanen Normalmodenanalyse (INMA)" beschrieben. Dabei wird durch eine MM Molekuldynamik-(MD-)Simulation ein Raumtemperatur-Ensemble von Strukturen des Moleküls in seiner jeweiligen Umgebung generiert und für jeden solchen Strukturschnappschuss das Infrarot-(IR-)Spektrum durch DFT/MM Normalmodenanalyse berechnet. Das inhomogen verbreiterte Raumtemperaturspektrum ergibt sich durch Überlagerung dieser Linienspektren.

Für BR ist die Erzeugung eines solchen Ensembles schwierig, da die BR Struktur bei Verwendung der üblichen nicht-polarisierbaren MM-Kraftfelder unter MD "zerfällt". Einen Ausweg bietet ein mit DFT/MM Methoden berechnetes polarisiertes BR-Kraftfeld, das die BR-Struktur auch bei Langzeit-MD-Simulationen (50 ns) erhält. Mit einem speziell entwickelten Hamiltonschen Replika-Austauschverfahren lässt sich anschließend zeigen, dass die bei den tiefen Temperaturen der Kristallstrukturanalyse homogene BR Struktur bei Raumtemperatur heterogen wird, was eine Vielzahl experimenteller Befunde erklärt. Die INMA Berechnung der Schwingungsspektren des BR-Chromophors bei Raumtemperatur *in situ*, eine systematische Analyse diverser Einflussparameter und Vergleiche mit experimentellen Daten beantworten schließlich die Frage, mit welcher Qualität solche Spektren bei sorgfältiger Modellierung und unter Einsatz der gewöhnlich verfügbaren DFT/MM Methoden berechnet werden können.

Weiterhin werden DFT/MM-Hybridrechnungen dazu eingesetzt, um den strukturellen Gehalt von ps-zeitaufgelöste IR Daten zur Entfaltungsdynamik eines lichtschaltbaren  $\beta$ -Hairpinpeptids zu dekodieren. Der lichtinduzierte Entfaltungsprozess wird mit MM/MD simuliert. DFT/MM Berechnungen der durch den Entfaltungsprozess

ausgelösten Bandenverschiebungen im Amid-I Bereich zeigen dann, dass die spektroskopischen Beobachtungen das sukzessive Aufreissen der ursprünglich geordneten H-Brückenstruktur des Peptids anzeigen.

# Abkürzungsverzeichnis

ATP	Adenosintriphosphat 5
BR	Bakteriorhodopsin 5
DFT	Dichtefunktionaltheorie 19
DNA	Desoxyribonukleinsäure 2
FTTCF	Fourier transform of the time correlation function 21
INMA	Instantane Normalmodenanalyse 21
IR	Infrarot 26
MD	Molekulardynamik 10
MM	Molekülmechanisch 10
NMR	Nuclear Magnetic Resonance 15
PR	Proteorhodopsin 7
RSBH <sup>+</sup>	protonierte Schiffische Base des Retinals 5
RR	Resonanz-Raman 18
TF-GV	Zeitaufgelöstes, fehlerbereinigtes generalisiertes Virial 21



# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>iii</b>
<b>Abkürzungsverzeichnis</b>	<b>v</b>
<b>1 Einleitung</b>	<b>2</b>
1.1 Struktur von Proteinen . . . . .	2
1.1.1 Bakteriorhodopsin . . . . .	5
1.1.2 $\beta$ -Hairpinpeptid . . . . .	11
1.2 Experimentelle Methoden . . . . .	13
1.2.1 Röntgenbeugung . . . . .	13
1.2.2 NMR-Spektroskopie . . . . .	15
1.2.3 Schwingungsspektroskopie . . . . .	15
1.3 Theoretische Methoden und Beschreibungen . . . . .	19
1.3.1 Methoden zur Berechnung von IR Spektren . . . . .	19
1.3.2 MM-MD-Simulationen . . . . .	22
1.4 Ziele und Gliederung dieser Arbeit . . . . .	25
<b>2 Berechnung der IR-Spektren des Retinalchromophors von BR</b>	<b>29</b>
2.1 Stabilisierung der Struktur des Chromophorbindungstasche in MD-Simulationen durch Berücksichtigung von Polarisierungseffekten . . . .	29
2.2 DFT/MM-Schwingungsanalyse des Retinalchromophors von BR . . .	59
<b>3 Untersuchung der Konformationsdynamik eines <math>\beta</math>-Hairpinpeptides</b>	<b>89</b>
<b>4 Zusammenfassung und Ausblick</b>	<b>103</b>
<b>Literaturverzeichnis</b>	<b>111</b>
<b>Danksagung</b>	<b>119</b>
<b>Lebenslauf</b>	<b>121</b>



*„Denn auf Mischung kommt es an.“*

Johann Wolfgang von Goethe

# 1 Einleitung

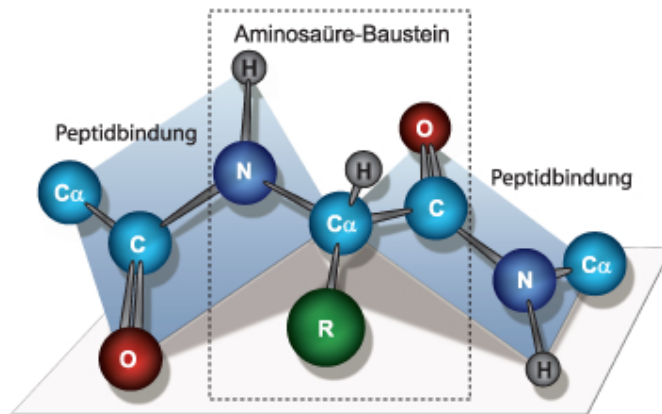
In allen lebenden Zellen sind Proteine die wichtigsten Funktionsträger zellulärer und physiologischer Prozesse. Aus Genomanalysen [4] kann man abschätzen, dass eine menschliche Zelle die Information zur Synthese von etwa 20000–25000 Proteinen besitzt. So groß wie diese Zahl ist auch das Aufgabenfeld der Proteine in der Zelle und ihre strukturelle Vielfalt, die sie benötigen, um ihre diversen Aufgaben zu bewältigen. Im Hinblick auf die Frage nach ihrer Funktion ist es außerordentlich wichtig zu wissen, wie diese Biomoleküle auf atomarer Ebene aufgebaut sind. Auf den nächsten Seiten möchte ich deswegen einen kurzen Überblick über die Proteinarchitektur geben und die Biomoleküle vorstellen, die in der vorliegenden Arbeit untersucht werden.

## 1.1 Struktur von Proteinen

Proteine bestehen aus 20 natürlichen Aminosäuren, deren Aufeinanderfolge in der Erbsubstanz Desoxyribonukleinsäure (DNA, vom engl. deoxyribonucleic acid) kodiert ist [5]. Jede Aminosäure setzt sich aus immer den gleichen Bauteilen zusammen: einem zentralem Kohlenstoffatom ( $C_\alpha$ ), einer Carboxylgruppe ( $\text{COOH}$ ), einer Aminogruppe ( $-\text{NH}_2$ ), einem Wasserstoffatom und einem Molekülrest ( $R$ ), der so genannten *Seitengruppe*. In dieser Seitengruppe liegt der einzige Unterschied der Aminosäuren [6]. So gibt es hydrophile und hydrophobe, saure und basische, aromatische und aliphatische Seitengruppen. Die Verknüpfung der diversen Aminosäuren erfolgt dabei durch Wasserabspaltung und Ausbildung torsionsstabiler Peptidbindungen  $\text{CO}-\text{NH}$  (siehe Abbildung 1.1).

Im Gegensatz zu dem DNA-Molekül, das in der Zelle als Doppelhelix vorliegt [5], können Proteinmoleküle die verschiedensten dreidimensionalen Strukturen ausbilden. Die spezifischen Strukturen, in die sie sich nach der Synthese falten [7], bestimmen ihre unterschiedlichen Funktionen in der Zelle [8]: Proteine beschleunigen als Enzyme chemische Reaktionsraten um viele Größenordnungen, wandeln Licht in die chemische Energie um, dienen als molekulare Motoren, replizieren die DNA-Helix, etc. Die Proteinstrukturen sind äußerst komplex und lassen sich auf drei Hierarchieebenen beschreiben:

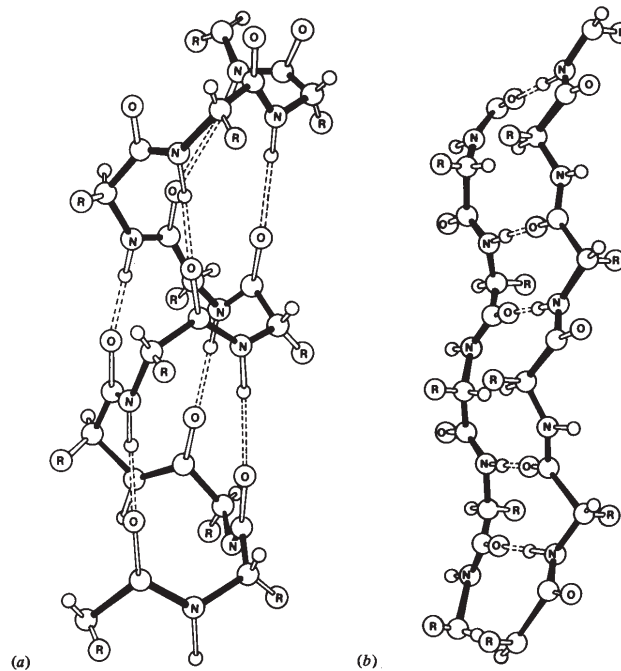




**Abbildung 1.1:** Schematische Darstellung eines Aminosäure-Bausteines in einem Polypeptid. Die Peptidbindungen sind als starre Flächen dargestellt, da sie relativ torsionsstabil sind. Konformationsänderungen eines Proteins entstehen durch Drehungen um die Bindungen der zentralen  $C_{\alpha}$ -Atome.

Die Primärstruktur oder Aminosäuresequenz beschreibt die Reihenfolge der Aminosäuren in der eindimensional gestreckten Polypeptidkette, wie sie die in der DNA gespeicherte Information vorgibt. In dieser langgestreckten Form liegen Proteine allerdings nie vor. Schon unmittelbar nach der Synthese bilden bestimmte kurze Sequenzen von Aminosäuren innerhalb von wenigen Mikrosekunden [9] lokale räumliche Anordnungen, die sog. Sekundärstrukturelemente.

Wichtige Beispiele hierfür sind die  $\alpha$ -Helices und die  $\beta$ -Faltblätter. In Abbildung 1.2 (a) ist ein kleiner Ausschnitt aus einer  $\alpha$ -Helix dargestellt. Das Rückgrat der  $\alpha$ -Helix hat die Form einer Spirale, die durch Wasserstoffbrücken stabilisiert wird. Dabei sind die Peptidgruppen so angeordnet, dass ihre elektrischen Dipole parallel hintereinander liegen. Die Superposition aller Dipole einer  $\alpha$ -Helix ergibt einen Makrodipol. Um die elektrostatische Wechselwirkung dieser Makrodipole zu optimieren, lagern sich  $\alpha$ -Helices gerne antiparallel aneinander. In Abbildung 1.2 (b) ist ein weiteres Beispiel für die Sekundärstruktur, ein  $\beta$ -Faltblatt, zu sehen. Zwei oder mehr nebeneinander liegende Abschnitte einer Polypeptidkette werden durch Wasserstoffbrücken zwischen den dipolaren NH- und CO-Gruppen des Peptids zusammengehalten. Die Dipole aufeinanderfolgender Peptidgruppen sind hier antiparallel nebeneinander ausgerichtet, was der  $\beta$ -Faltblatt-Struktur eine hohe Stabilität verleiht. Das Rückgrat des Proteins macht dabei bei jedem zentralen C-Atom einen Knick und beschreibt so eine Zickzacklinie. Dadurch entsteht eine Struktur, die an ein gefaltetes Blatt erinnert. Die Strukturelemente fügen sich im Protein zu einer komplexen, jedoch wohl definierten räumlichen Struktur, zur Tertiärstruktur des Proteins, zusammen, wobei die verschiedenen Sekundärstrukturelemente durch sogenannte Schleifen (engl.



**Abbildung 1.2:** Sekundärstrukturen in Proteinen. Das Proteinrückgrat ist schwarz, Wasserstoffbrücken sind gestrichelt gezeichnet. (a) Ausschnitt aus einer  $\alpha$ -Helix. Das Rückgrat der Helix hat die Form einer Spirale, die durch Wasserstoffbrücken stabilisiert wird. (b) Ausschnitt aus einer  $\beta$ -Faltblatt-Struktur. Zwei nebeneinander liegende Abschnitte einer Polypeptidkette sind durch Wasserstoffbrücken miteinander verbunden. Das Rückgrat des Proteins macht dabei bei jedem zentralen C-Atom einen Knick, so dass eine Struktur entsteht, die an ein gefaltetes Blatt erinnert. Entnommen aus [10].

Loops) miteinander verbunden sind. Im Gegensatz zu den rigiden, festkörperartigen  $\alpha$ -Helices und  $\beta$ -Faltblättern sind Schleifen häufig relativ flexibel und ermöglichen so funktional wichtige Prozesse der Konformationsdynamik, bei denen ein Protein seine Tertiärstruktur ändern kann.

Die korrekte dreidimensionale Struktur ist eine Voraussetzung für die korrekte Funktion eines Proteins in der Zelle. Fehler bei der Faltung führen dazu, dass Proteine die für sie genetisch bestimmten Aufgaben nicht übernehmen können, was für die Zelle schwerwiegenden Folgen hat [11].

### 1.1.1 Bakteriorhodopsin

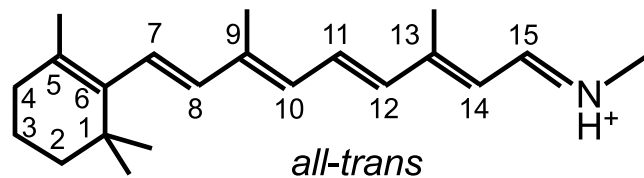
Ich möchte mich nun jenem Protein zuwenden, dessen Untersuchung im Zentrum der vorliegenden Arbeit steht. Es handelt sich dabei um das Retinalprotein Bakteriorhodopsin (BR). Das BR ist zur Zeit das wohl am besten erforschte photoreaktive Protein [12–21]. Es ist in die Membran des Archaeobakteriums *Halobacterium Salinarium* eingebaut, das in extrem salzigen Gewässern vorkommt (siehe Abbildung 1.3). Wie Oesterhelt und Stoeckenius gezeigt haben [22], ist das BR eine lichtgetriebene Protonenpumpe, die Protonen aus der Zelle transportiert. Der aufgebaute Protonengradient zwischen der extra -und intrazellulärem Raum kann anschließend vom Bakterium zur Synthese von Adenosintriphosphat (ATP), dem allgemeinen zellulären Treibstoff, genutzt werden [22].



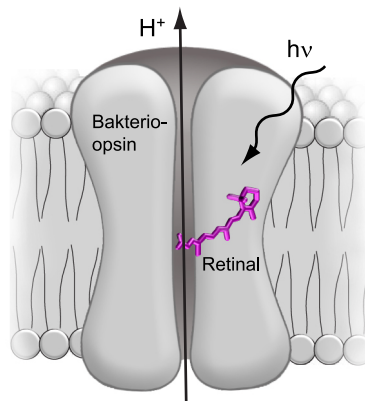
**Abbildung 1.3:** Halobakterien, die das Bakteriorhodopsin in der Membran enthalten, findet man in extrem salzigen Seen. Durch die Lichtabsorption im grünen Spektralbereich, verleiht es solchen Gewässern eine purpurne Farbe. Bild entnommen aus [23]

Bakteriorhodopsin ist ein tiefpurpurfarbener Farbstoff-Protein-Komplex [24]. Es besteht aus 248 Aminosäuren, deren Abfolge aus Aminosäuren- und DNA-Sequenzierungsstudien [25] bekannt ist, und dem Farbstoff Retinal, der Licht aus dem visuellen Spektralbereich absorbiert. Die Bindung des Retinals an das Protein erfolgt kovalent als eine protonierte Schiff'sche Base ( $\text{RSBH}^+$ ) an einen Lysinrest. Abbildung 1.4 zeigt die chemische Struktur der  $\text{RSBH}^+$  in ihrer gestreckten *all-trans*-Geometrie. Die  $\text{RSBH}^+$  besteht aus einer Polyenkette konjugierter Doppelbindungen von C-Atomen und einem Zyklhexenring. Nach der Lichtabsorption ändert das Retinal seine Geometrie, gibt das Proton der Schiff'schen Base an eine geladene Aminosäure (Asp85)

in seiner Umgebung ab und setzt damit eine Reaktion im Gang [26], in deren Folge ein Proton aus der Zelle transportiert wird (siehe Abb. 1.5). Nach dem Transport eines Protons kehrt das Protein wieder in seinen Ausgangszustand zurück. Es benötigt dafür keine weiteren Kofaktoren und funktioniert also zyklisch.



**Abbildung 1.4:** Chemische Struktur der protonierten Schiff'schen Base des Retinal ( $\text{RSBH}^+$ ) in der *all-trans*-Geometrie.



**Abbildung 1.5:** BR besteht aus dem Protein Opsin und dem Farbstoff Retinal, der als  $\text{RSBH}^+$  kovalent an das Protein gebunden ist. Die  $\text{RSBH}^+$  absorbiert Licht aus dem visuellen Spektralbereich und initiiert einen vektoriellen Protonentransport über die Membran aus dem Zytoplasma (unten) in den Zellaußenraum (oben). Der dadurch erzeugte Protonengradient kann anschließend von der Zelle zur ATP-Synthese verwendet werden.

Nur bei wenigen anderen Proteinen sind vergleichbar viele Details der Struktur und Funktionsweise wie bei BR aufgeklärt worden. Das besondere Interesse an BR hatte die folgenden Gründe. Erstens ist es wahrscheinlich das einfachste biologische System, das Licht in chemische Energie umwandeln kann. Zweitens ist es außerordentlich stabil, kann leicht erzeugt werden und durchläuft nach der Photoanregung

eine zyklische Reaktion, so dass anhand des BR zahlreiche moderne Techniken der experimentellen Biophysik entwickelt und erprobt werden konnten [20, 21, 27–29].

Drittens ist BR ein Protein, das dem Photorezeptor in den Sehzellen der Wirbeltiere, Rhodopsin, sehr ähnlich ist [22]. Wie auch BR besteht Rhodopsin aus einem Apoprotein und einem kovalent an einen Lysinrest gebundenen Retinalchromophor. In der Retina vorkommend dient Rhodopsin jedoch nicht der Energiewandlung sondern der Umwandlung von Licht wahrgenommene Bilder, die in Form paralleler elektrochemischer Signale über den Sehnerv an das zentrale Nervensystem weitergeleitet werden. Nach Lichtabsorption und Initiierung eines solchen Signals zerfällt das Rhodopsin irreversibel zu Opsin und zu freiem Retinal [30], was die experimentelle Aufklärung seiner molekularen Struktur und der Funktionsweise außerordentlich schwierig macht. Im Gegensatz dazu bleibt BR jedoch nach Lichtabsorption stabil, so dass sich Untersuchungen am BR einfacher durchführen lassen.

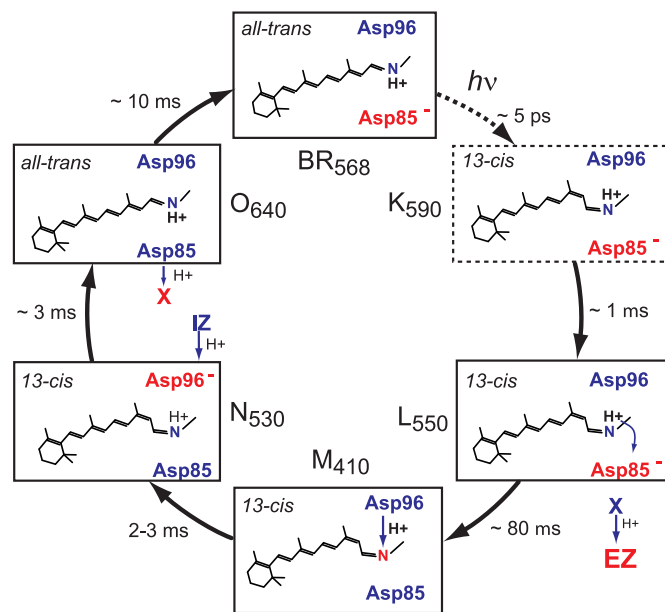
Bis vor einigen Jahren galt die Annahme, dass die Retinalproteine BR, sowie die verwandte Chloridionenpumpe *Halorhodopsin* und die ebenfalls verwandten sensorischen Rhodopsine SRI und SRII, die allesamt in *Halobacterium Salinarium* vorkommen, eine isolierte Proteinfamilie darstellen. Doch in den letzten fünf Jahren wurde noch eine Vielzahl verwandter Opsine beispielsweise in *E. coli* oder im Meeresplankton (das *Proteorhodopsin*) entdeckt [31, 32]. Sowohl bei den Opsinen aus *E. coli* als auch bei dem Proteorhodopsin führt die Lichtabsorption zu einer ähnlichen Photoreaktion wie bei BR. Mittlerweile scheint es also, als ob diese Proteine einen früher unbekannten Weg für die Nutzung der Lichtenergie in der Ozeanwelt vermitteln und die Familie der Retinalproteinen folglich eine sehr wichtige Komponente der Erdbiosphäre ist [33].

Aufgrund seiner Stabilität und leichter Gewinnbarkeit wurden am BR zahlreiche Experimente [28, 34–36] durchgeführt, die die Frage klären sollten, auf welche Weise Protonenpumpen Protonen durch eine Membran transportieren. Der Prozess der Protonentransfer dauert bei BR ca. 10 ms [37]. Dabei bewegen sich Protonen im BR durch Sprünge von Donor- zu Akzeptorgruppen, welche sich im Protein nah an einander befinden. Schon lange vor der Aufklärung der BR-Struktur hatten die am Chromophor sowohl im optischen als auch im infraroten Spektralbereich beobachteten starken Solvatochromieeffekte (die sog. *Opsin-Verschiebungen*) auf die Existenz negativ geladener Peptidreste Asp85, Asp212 in der Retinalbindungstasche des BR hingewiesen [34, 38].

Dieses Ergebnis steht im Einklang mit den Arbeiten von Siebert et al. [28] und Braiman et al. [36], die anhand Fourier-Transform-Infrarot-(FTIR-)Untersuchungen an spezifischen BR-Mutanten zeigten, dass die Aspartate Asp85 und Asp96 am Pro-

tonentransport beteiligt sind. Die Änderungen von Protonierungszuständen dieser Gruppen während des Protonentransports kann man mit FTIR Methoden spektroskopisch identifizieren und Intermediaten K, L, M, O des sog. *Photozyklus* zuordnen.

Abb.1.6 zeigt den Ablauf des zunächst durch optische Absorptionsspektroskopie charakterisierten Photozyklus von BR. Das initiale Ereignis ist die Absorption eines Lichtquants durch die RSBH<sup>+</sup>. Im Ausgangszustand liegt die RSBH<sup>+</sup> in der gestreckten *all-trans* Konfiguration vor (siehe Abb. 1.4) und ihr Absorptionsmaximum liegt dann bei 568 nm. Einige Pikosekunden nach der Lichtanregung bildet sich das K-Intermediat [39], in dem das Retinal in die 13-*cis*-Konfiguration isomerisiert ist und ein rotverschobenes Absorptionsmaximum hat [40, 41]. Dabei ist die Chromophor-geometrie aber zunächst stark verspannt und liegt erst nach ca. 1  $\mu$ s in einer etwas entspannteren 13-*cis*-Konfiguration (L-Intermediat) vor, was sich in einer Blauverschiebung des Absorptionsmaximums nach 550 nm zeigt [42]. Bei dem Zerfall des



**Abbildung 1.6:** Photozyklus von BR nach [37]. Für das jeweilige Intermediat sind die Geometrie der RSB, ihr Absorptionsmaximum und die Protonierungszustände der für den Protonentransport wichtigen Gruppen gezeigt. Die Protonierungszustände sind farblich kodiert: protonierte Gruppen sind blau und deprotonierte Gruppen sind rot gezeichnet.

L-Intermediats gibt die Schiffische Base ihr Proton an das Asp85 ab [36, 43]. Diese Protonübertragung ist erst durch die vorherige Isomerisierung des Retinals möglich,

denn im Ausgangszustand besitzt die RSB einen  $pK_S$ -Wert von über 13 und liegt demzufolge unter physiologischen Bedingungen protoniert vor. Der Protonenakzeptor Asp85 besitzt dagegen einen  $pK_S$ -Wert von etwa 4, ist also deprotoniert.

Nach einem Vorschlag von Schulten und Tavan [40] führt die lichtangeregte Isomerisierung im L-Intermediat zu einer um Einfachbindungen der Polyenkette verdrillte Geometrie der RSBH<sup>+</sup>. Solche Bindungstorsionen können, wie quantenchemische Rechnungen gezeigt haben, den  $pK_S$ -Wert des Chromophors so stark erniedrigen, dass die RSBH<sup>+</sup> ihr Proton an Asp85 abgeben kann [40]. Dieser Vorschlag war aber keineswegs allgemein akzeptiert [19, 42, 44, 45]. Der Hauptgrund hierfür war, dass FTIR Spektren eines durch Abkühlen auf 150 K eingefrorenen "L-Intermediates" auf eine planare Chromophorstruktur hingewiesen hatten. Eine der Motivationen der vorliegenden Arbeit war, dass spätere (am L Intermediat des physiologischen Pumpzyklus durchgeführte) zeitaufgelöste Messungen verbleibende Verdrillungen der Polyenkette nachgewiesen haben [46], was eine eingehendere Analyse dieser Spektren notwendig erscheinen ließ.

Durch die Protonenabgabe der RSB beim Übergang vom L- zum M-Intermediat [47], ca. 80  $\mu$ s nach der Lichtabsorption, wird ihr Absorptionsspektrum ins Blaue verschoben [48]. Zugleich gibt eine noch nicht genau bekannte Gruppe (in der Abbildung 1.6 ist sie als X gekennzeichnet) aus Aminosäuren und Wassermolekülen ein Proton an die extrazelluläre Seite ab.

Die Reprotonierung der Schiffschen Base führt zum N-Intermediat. Der Protonendonator für die Schiffsche Base ist das Asp96 [49]. Das Asp96 hat in BR aufgrund seiner hydrophoben Umgebung zunächst einen sehr hohen  $pK_S$ -Wert von über 11 [50, 51]. Es liegt also im Ausgangszustand protoniert vor. Das Asp96 befindet sich auf der dem Zytoplasma zugewandten Seite ca. 12 Å von der Schiffschen Base entfernt. Der Protonentransfer von Asp96 zur Schiffschen Base über diese Distanz ist nur durch die Bildung eines transienten Wasserstoffbrückennetzes aus einströmenden Wassermolekülen zu erklären [52], welche durch eine Konformationsänderung des Proteins ermöglicht wird.

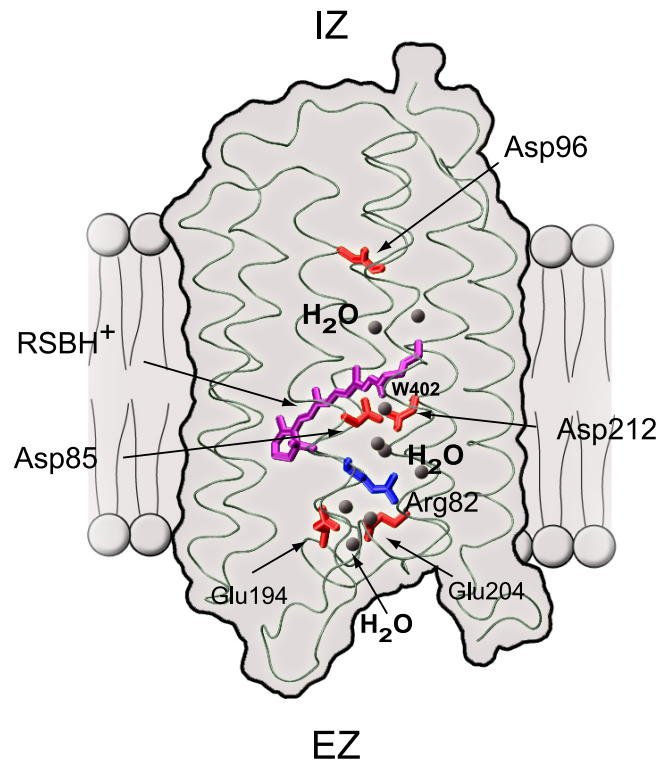
Beim N-Zerfall wird Asp96 vom Zytoplasma aus reprotoniert und die RSBH<sup>+</sup> isomerisiert thermisch in die *all-trans*-Geometrie zurück [53, 54]. Der O-Zerfall ist schließlich durch die Relaxation der Konformationsänderungen im Protein [55] und den Übergang des Protons von Asp85 zu der an der extrazellulären Seite liegenden Aminosäurenggruppe charakterisiert, so dass nach 10 ms der Ausgangszustand wieder hergestellt ist.

Wie man sieht, haben spektroskopische Untersuchungen am BR [27, 29, 35, 36, 39–41, 43, 47, 48, 53] eine Vielzahl von Informationen über seinen Photozyklus und die Geometrie des Chromophors geliefert. Andererseits war und ist die strukturelle Interpretation dieser Spektren [19, 29, 42, 45] so schwierig, dass in einer Reihe von Fragestellungen, beispielsweise hinsichtlich der Geometrie des Chromophors im L-Intermediat des BR Photozyklus, keine eindeutigen Antworten erzielt werden konnten. Die in [45] dokumentierte Debatte legt hierfür lebendiges Zeugnis ab. Neu entwickelte Hybridverfahren [56], welche es gestatten, die DFT-Beschreibung eines Chromophors in ein molekülmechanisches (MM) Modell seiner Umgebung einzubetten, und dazu entwickelte Rechenprotokolle [57] erlauben mittlerweile eine viel genauere Beschreibung der Schwingungsspektren des Chromophors. Wir werden in der vorliegenden Arbeit ausloten, welche Genauigkeit bei solchen Beschreibungen erzielt werden kann [1, 2]. Eine wesentliche Voraussetzung für die Anwendung der angesprochenen Methoden ist jedoch die detaillierte Kenntnis der zugehörigen Proteinstruktur.

Glücklicherweise ist die Struktur des Ausgangszustands von BR mittlerweile anhand von Röntgenbeugung-Experimenten sehr gut geklärt [15]. Die in Abbildung 1.7 visualisierte kristallographische Struktur 1C3W [15] wurde mit einer hervorragenden Auflösung von 1.55 Å bei kryogenen Temperaturen bestimmt. Die kristallographischen Untersuchungen zeigen, dass manche funktionell wichtige Gruppen derart weit voneinander entfernt sind, dass direkter Wasserstoffbrückenkontakt und damit Protonenaustausch zwischen ihnen nicht möglich ist. Neben diesen Residuen, ist also die Gegenwart von Wassermolekülen entscheidend für die Funktionsfähigkeit von BR [58]. Insbesondere ist ein solches Wassermolekül (W402) direkt an die RSBH<sup>+</sup>, die zentrale Protonenbindestelle des Proteins, angelagert. Erst durch die internen Wassermoleküle wird eine Verbindung zwischen den Residuen geschaffen, so dass Protonentransfer möglich wird.

Wie auch die Untersuchungen zum BR gezeigt haben, die im Verlaufe der vorliegenden Arbeit durchgeführt wurden und auf die möglichst genaue Beschreibung der Chromophorspektren zielten, ist selbst ein so relativ kleines Protein wie das BR noch so komplex, dass es sich einer Simulationsbeschreibung, auch bei Einsatz modernster Methoden zum Abtasten von Konfigurationsräumen, weitgehend entzieht. So dauert der Photozyklus des BR, wie wir oben gesehen haben, etwa 10 ms, während die Simulationszeiten, die man zu dem heutigen Stand der Technik mit Molekulardynamik-(MD-)Simulationen überdecken kann, auch bei deutlich kleineren Proteinen höchstens im Bereich von einigen wenigen Mikrosekunden liegen [59]. Um für lichtinduzierte dynamische Prozesse in Polypeptiden MD-Beschreibungen mit experimentellen Daten





**Abbildung 1.7:** Extrazellulärer Protonentransportkanal [15] mit funktionell wichtigen Gruppen und den Positionen einiger interner Wassermoleküle. Bakteriorhodopsin in die Purpurmembran des *Halobacterium Salinarium* eingebaut. BR pumpt Protonen vom Zellinneren (oben) in den extrazellulären Raum (unten).

vergleichen zu können, ist also eine Reduktion der Komplexität der zu untersuchenden Systeme dringend nötig.

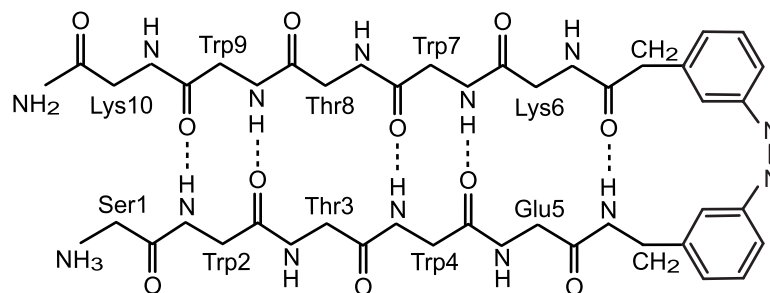
### 1.1.2 $\beta$ -Hairpinpeptid

Um Theorie und Experiment näher aneinander zu bringen, wurde daher im Rahmen des SFB533 "Lichtinduzierte Dynamik in Biopolymeren" ein Projekt gestartet, das ein sehr viel kleineres System in den Untersuchungsfokus nahm.

Es handelte sich dabei um ein  $\beta$ -Hairpinpeptid, das ein Modell für eine  $\beta$ -Faltblatt-Struktur darstellt. In das gewählte Modellpeptid, dessen chemische Struktur in Abbildung 1.8 dargestellt ist, wurde ein Azobenzolfarbstoff als lichtschaltbare Gruppe eingebaut. In der gezeichneten *cis*-Konfiguration modelliert der Azobenzolschalter sehr gut die Stereochemie eines  $\beta$ -Turns in einem nativen Hairpin. Nach lichtinduzierter

*cis-trans*-Isomerisierung erzwingt der Azobenzolschalter eine Konformationsänderung des Peptids, die anscheinend auf einer Zeitskala von wenigen Nanosekunden abgeschlossen ist [60]. Das skizzierte Azopeptidmodell sollte daher sehr viel besser für den Vergleich experimenteller mit theoretischen Untersuchungen geeignet sein als das hochkomplexe und viel langsamer relaxierende BR.

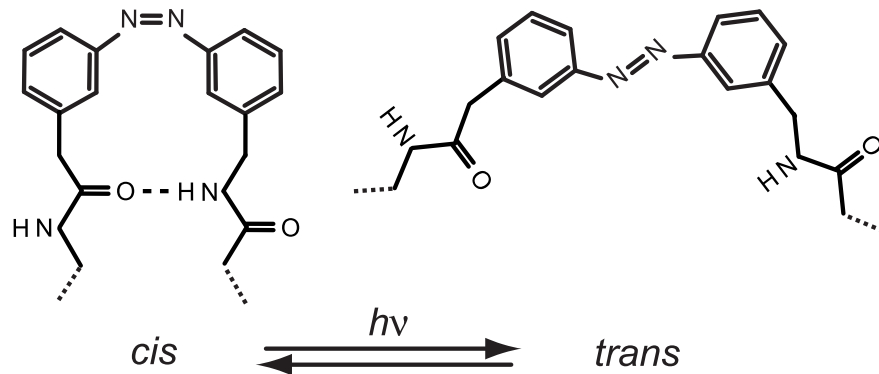
Abbildung 1.8 zeigt die synthetisierte Aminosäuresequenz  $\text{NH}_3\text{-Ser-Trp-Thr-Trp-Glu-AMPP-Lys-Trp-Thr-Trp-Lys-NH}_2$  [60], die zwei mit einem optischen Schalter, dem Azobenzolfarbstoff AMPP, verbundene Stränge ausbildet. Das Auftreten von



**Abbildung 1.8:** Aminosäuresequenz und chemische Struktur des untersuchten  $\beta$ -Hairpins. Rechts ist der Schalter Azobenzol zu sehen, dessen *cis/trans* Isomerisierung die Entfaltung/Faltung auslöst.

zwei stabilen Konformationen des Azobenzolmoleküls ist seine bedeutendste Eigenschaft. Wie die Abbildung 1.9 zeigt, werden die beide Strukturen durch Isomerisierung der  $\text{N}=\text{N}$ -Doppelbindung gebildet und als *cis*- bzw. *trans*-Konformation bezeichnet. Die Umwandlung beider Konformere ineinander kann photochemisch durch Belichten mit blauem Licht unterschiedlicher Wellenlänge gesteuert werden. Die kompaktere *cis*-Form hat eine Länge von ca. 7 Å. Zwei Peptid-Stränge können damit Wasserstoffbrücken zu einander ausbilden, so dass eine spezifische  $\beta$ -Hairpin-Struktur entsteht. Das Belichten mit Laserpulsen bei 420 nm führt zu einer Streckung des Azobenzolmoleküls um etwa 2 Å.

Wie in der vorliegenden Arbeit durch Vergleich berechneter mit beobachteten zeitaufgelösten Schwingungsspektren gezeigt werden wird [3], bewirkt die strukturelle Änderung des Schaltermoleküls, dass die in seiner Nähe liegende Wasserstoffbrücken zwischen den beiden Peptidsträngen brechen und eine Entfaltungsreaktion in einer Art Reißverschluss initiiert wird. Wie gezeigt werden wird, verläuft die Reaktion über einige wenige Intermediate und ist innerhalb von 2 ns mit einem resultierendem Ensemble geknäulter Strukturen abgeschlossen [3]. Die umgekehrte Reaktion erfordert dagegen, dass die Peptidstränge sich durch Wasserstoffbrücken stabilisieren, was viel



**Abbildung 1.9:** Modell für *cis*- bzw. *trans*-Isomerisierung des Azobenzolmoleküls. Zusätzlich ist einen Ausschnitt aus der  $\beta$ -Hairpin Peptidstruktur abgebildet. Wasserstoffbrücken zwischen den Peptidsträngen können nur in der *cis*-Konformation ausgebildet werden. Die Isomerisierung des Photoschalters in die *trans*-Konformation führt zur Entfaltung des Peptides.

mehr Zeit in den Anspruch nimmt, so dass die Faltungsreaktion erst in Bereich von  $30 \mu\text{s}$  erfolgen kann [60].

## 1.2 Experimentelle Methoden

Wie man am Beispiel des BR sehen kann, werden für die Untersuchungen von dynamischen Prozessen in Proteinen zahlreiche Methoden der experimentellen Biophysik benötigt. Da wir die Ergebnisse entsprechender Experimente für unsere eigenen theoretischen Modellbildungen sowohl verwenden als auch kritisch diskutieren werden, scheint an dieser Stelle eine kurze Diskussion dieser Methoden angebracht zu sein.

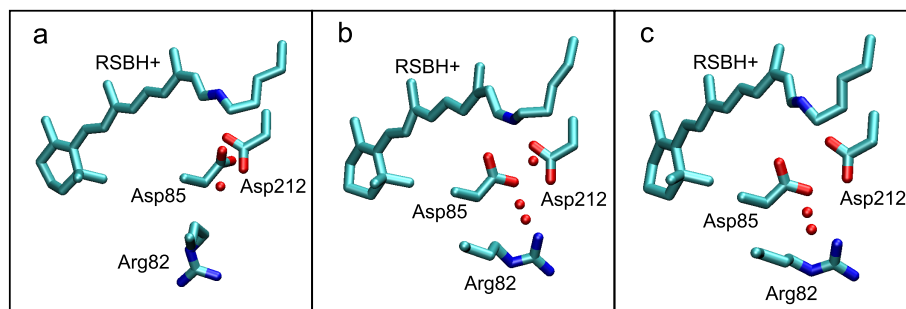
### 1.2.1 Röntgenbeugung

Gegenwärtig ist die Röntgenbeugung die wichtigste Methode, um die Struktur eines Proteins mit atomarer Auflösung zu bestimmen. Grundvoraussetzung der Methode ist die Verfügbarkeit eines gut geordneten Proteinkristalls [61, 62]. Die Kristallisation speziell von Membranproteinen in einer Weise, dass die native Struktur des zu kristallisierenden Moleküls möglichst wenig gestört wird und gleichzeitig gut geordnete Kristalle entstehen, ist eine weitgehend alchimistische Kunst, für die es bislang kaum empirische Regeln gibt. Deshalb weisen solche Proteinkristalle häufig Defekte unter-

schiedlicher Art auf, welche die Konstruktion einer verlässlichen atomaren Struktur aus Beugungsdaten sehr erschweren oder gar verhindern können. Für den Fall von BR hat Lanyi [63, 64] all jene experimentellen Probleme eingehend erörtert, welche sich dem Versuch entgegenstellen, den Ausgangszustand und die Intermediate mit Röntgenbeugungsmethoden genau zu erfassen.

Eines dieser Probleme sind die Strahlenschäden an der Kristallstruktur, welche durch harte Röntgenstrahlung verursacht werden. Die Strahlung erzeugt freie Radikale, die sich bei Raumtemperatur schnell durch den Proteinkristall bewegen können und das Gitter durch chemische Reaktionen zerstören [65]. Um die Kristallschädigung bei der Strukturanalyse gering zu halten, werden Kryotechniken [66, 67] angewandt, d.h. die Messungen werden bei tiefen Temperaturen um 100 K durchgeführt. Dabei ist allerdings nicht klar, in welchem Maße die Tieftemperaturstrukturen mit den Strukturen bei physiologischen Temperaturen vergleichbar sind.

Ein konkretes Beispiel für die geschilderten Schwierigkeiten ist das L-Intermediat des BR. Abbildung 1.10 zeigt drei von verschiedenen Arbeitsgruppen vorgeschlagene Kristallstrukturen [68–70]. Wie aus der Abbildung zu erkennen ist, unterscheiden sich die drei Strukturen drastisch [63]. Zusätzlich zeigt die Abbildung 1.10 ein weiteres Problem bei der Proteinkristallographie. Wegen der kleinen Streuquerschnitte ist es nicht möglich, die Positionen der Wasserstoffatome in Proteinen zu bestimmen. Die Information über die Protonierungszustände basischer bzw. saurer Seitengruppen geht somit verloren, so dass beispielsweise das Verständnis des Protonentransferprozesses im BR allein anhand der kristallographischen Daten nicht möglich ist.



**Abbildung 1.10:** Visualisierte Kristallstrukturen des BR im L-Intermediat in der Umgebung des Chromophores: a) 1EOP [68], b) 1O0A [69], c) 1UCQ [70].

Im Vergleich zu den Intermediaten ist der Grundzustand des BR sehr gut charakterisiert. 1998 haben Luecke et al. [15] die Struktur von BR mit einer Auflösung von 2.4 Å bestimmt und sie 1999 bis zur Auflösung von 1.5 Å verfeinert (siehe Abbildung 1.7). Diese Auflösung entspricht etwa der Größenordnung von Bindungsabständen.

Damit konnte man also die einzelnen schweren Atome erkennen und die Seitenketten identifizieren.

### 1.2.2 NMR-Spektroskopie

Neben der Röntgenbeugung ist die multidimensionale NMR-Spektroskopie (Nuclear Magnetic Resonance) die wichtigste Methode, um Strukturen von Proteinen zu bestimmen [71]. Der Vorteil der NMR-Spektroskopie ist, dass Proteine direkt in der Lösung analysiert werden können, und die erhaltenen dreidimensionale Strukturen Proteine unter nahezu physiologischen Bedingungen repräsentieren. Die Methode ist jedoch aufwändiger als Röntgenstrukturanalyse und nur auf relativ kleine Proteine anwendbar.

Die NMR liefert näherungsweise Abstandsinformationen für Kerne mit nichtverschwindendem Spin (wie beispielsweise Wasserstoffkerne). In der Regel reicht diese Abstandsinformation aber nicht aus, um die Struktur des Proteins zu bestimmen. Vielmehr müssen Methoden der MD-Simulation (siehe dazu Abschnitt 1.3.2) angewandt werden, bei denen ein MM Kraftfeld mit den NMR-Abstandsinformationen zu einer Hybridenergiefunktion kombiniert wird. Dabei wird durch simuliertes Abkühlen [72, 73] eine Vielzahl möglicher und mit den NMR Daten verträglicher Proteinstrukturen berechnet.

Als Ergebnis der Abkühlversuche erhält man ein Ensemble von Strukturen, die mehr oder minder große Gesamtenergien haben. Zum Beispiel zeigen Patzelt et al. [74] 11 Strukturen für die Retinalbindungstasche im BR, die sie mittels NMR-Technik bestimmt haben. Im Vergleich zur bereits erwähnten Kristallstruktur von Luecke et al. [15] weisen die Strukturen des NMR-Ensembles kleine aber signifikante Unterschiede auf, wie z.B. in der Konformation der Lys216-Kette, an die der Retinalchromophor gebunden ist. Wie man diese strukturellen Diskrepanzen erklären kann und ob die Unterschiede in der Lysinkette Auswirkungen auf die Struktur und damit die Eigenschaften des Retinals haben können, ist eine spannende Frage, die in der vorliegenden Arbeit untersucht werden soll. Diese Fragestellung bedarf jedoch den Einsatz weiterer Untersuchungsmethoden.

### 1.2.3 Schwingungsspektroskopie

Für das Verständnis biologischer Prozesse gewinnt die Schwingungs-Spektroskopie zunehmend an Bedeutung, insbesondere weil sie Aussagen über dynamische Vorgänge in biologischen Systemen erlaubt, die mit Methoden der Strukturaufklärung wie

Röntgenstrukturanalyse und NMR nicht erhalten werden. Sie liefert Informationen nicht nur über die Struktur des Proteinrückgrats, Geometrie und Protonierungszustands von Aminosäureseitenketten, sondern macht auch Einsichten in ihre zeitliche Veränderungen zugänglich.

Zur Untersuchung schnellster dynamischer Prozesse in Molekülen in Echtzeit wurde in den letzten Jahren eine Reihe neuer Techniken im Bereich der Laserspektroskopie entwickelt [75]. Mit dem Konzept der Anreg-Abtast-Spektroskopie [76] können photochemische Prozesse durch Anregung mit einem ultraschnellen Laserimpuls gezielt ausgelöst und anhand der Absorption oder die Streuung elektromagnetischer Strahlung im mittleren infraroten Spektralbereich (Infrarot- bzw. Raman-Spektroskopie) verfolgt werden [75]. Die auf diese Weise erhaltene Spektren zeigen Schwingungsmoden von Molekülen, deren typische Frequenzen im mittleren infraroten Spektralbereich liegen. Da ihre Positionen und Intensitäten sehr empfindlich von der Geometrie des Moleküls und derer Ladungsverteilung abhängt, und können Schwingungsspektren somit zur Untersuchung lichtinduzierter Reaktionen in Proteinen [77] oder auch zur Verfeinerung von kristallographischen Modellen herangezogen werden [78].

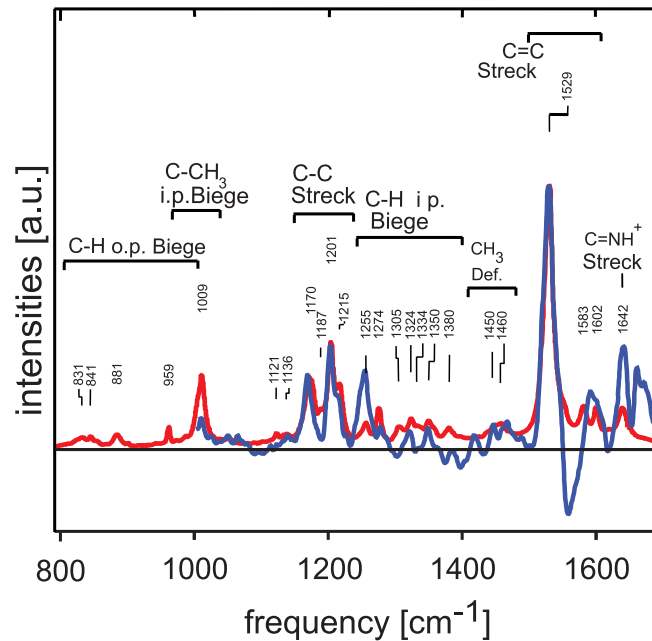
Proteine weisen mehrere charakteristische Schwingungsfrequenzen im infraroten Spektralbereich auf. Besonders interessant sind jedoch die so genannten Amid-Banden, die vom Proteinrückgrat hervorgerufen werden, da sie Rückschlüsse auf die Proteinsekundärstruktur erlauben. Eine davon ist die AmidI-Bande bei ca.  $1650\text{ cm}^{-1}$ , die hauptsächlich durch eine C=O Streckschwingung der Peptidbindung und geringen Anteilen der C–N Streckschwingung und der N-H Deformationsschwingung hervorgerufen wird. Da die Amidgruppen der Proteine an der Ausbildung von intramolekularen Wasserstoffbrückenbindungen beteiligt sind, die Sekundärstrukturelemente stabilisieren, hängt ihre Schwingungsfrequenz sehr empfindlich (siehe Tabelle 1.1) von der Art der ausgebildeten Sekundärstruktur ab und man kann durch eine Analyse der Bandelage Rückschlüsse auf die Sekundärstruktur des untersuchten Proteins ziehen.

In Makromolekülen, wie Proteinen, lassen sich auf der Basis von IR-Absorptionsspektren jedoch nur begrenzte Aussagen über die Strukturen innerhalb des Moleküls machen. Die hohe Anzahl der sich überlappenden IR-Banden ergibt einen breiten Absorptionshintergrund und macht eine genaue Zuordnung einzelner Schwingungen schwierig. Zur Unterscheidung einzelner, funktionell relevanter Atomgruppen wird daher die Methode der reaktionsinduzierten Differenzspektroskopie angewandt. Differenzspektren werden durch die Subtraktion von Absorptionsspektren zweier unterschiedlicher Zustände des Proteins, z.B. eines Photoprodukts und des entsprechenden Ausgangszustandes, erhalten. Dadurch wird der Großteil der sich nicht ändernden

Zuordnung	Bandenposition in H <sub>2</sub> O [cm <sup>-1</sup> ]
$\alpha$ -Helix	1648-1657
$\beta$ -Faltblatt	1623-1641
	1674-1695
Schleifen	1662-1686
Ungeordnet	1642-1657

**Tabelle 1.1:** AmidI-Banden verschiedener Sekundärstrukturelemente nach [79].

Absorptionsbeiträge eliminiert, d.h. nur diejenigen Schwingungen liefern einen Beitrag, die sich in beiden Proteinzuständen unterscheiden. Damit wird die Methode sensitiv auf kleinste Absorptionsänderungen [61].



**Abbildung 1.11:** Blau: FTIR-Differenzspektrum für BR im Grundzustand und in seinem M-Intermediat. Abbildung entnommen aus [80]. Rot: RR-Spektrum des Retinalchromophors entnommen aus [81]. Die Zuordnung der Banden zu einzelnen Teilen des Chromophors wurde ebenso aus [81] entnommen.

Die blaue Linie in der Abb. 1.11 zeigt ein derartiges FTIR-Differenzspektrum, bei dem das Spektrum des BR-Ausgangszustandes von jenem des M-Intermediates abgezogen ist [80]. Prinzipiell lassen sich die dort gezeigten Differenzbanden aufteilen in solche, die durch den Chromophor verursacht werden, und solche, die vom Protein herrühren [80]. Da das Protein im Bereich bis etwa  $1500\text{ cm}^{-1}$  kaum absorbiert und da der unprotonierte 13-*cis* Chromophor im M-Intermediat viel schwächer absorbiert als die *all-trans*-RSBH<sup>+</sup> im Ausgangszustand, sind die negativen Banden hauptsächlich der *all-trans*-RSBH<sup>+</sup> zuzuordnen.

Eine andere Methode, die bei Chromophorproteinen ergänzend zur FTIR-Spektroskopie angewandt wird und selektiv Informationen über Schwingungen der einzelnen Atomgruppen liefern kann, ist die Resonanz-Raman-(RR-)Spektroskopie [61]. Das Messprinzip hier beruht wie auch bei Raman-Spektroskopie auf der inelastischen Lichtstreuung, die Frequenz des eingestrahnten Lichts entspricht jedoch jener des ersten elektronischen Übergangs in der Gruppe, die typischerweise im sichtbaren bzw. im UV-Spektralbereich liegt. Man erhält damit viel größere Polarisierbarkeitsänderungen der Gruppe entlang der entsprechenden Schwingungsmode und damit auch Resonanzverstärkung der zugehörigen Banden. Ramanlinien von allen anderen Atomgruppen erscheinen im Bezug auf die resonant verstärkte Banden sehr viel kleiner und werden daher im Spektrum nicht beobachtet.

Mit der RR-Spektroskopie werden Chromophore untersucht, da sie elektromagnetischer Strahlung im sichtbaren bzw. im UV-Bereich absorbieren. Die rote Linie in der Abbildung 1.11 zeigt ein typisches RR-Spektrum des Retinalchromophors im BR im Ausgangszustand und eine Zuordnung der diversen Banden zu einzelnen Teilen des Chromophors.

Die verschiedenen Anregungsmechanismen für Schwingungen haben zur Folge, dass sich die IR- und Ramanspektren in den Intensitäten der einzelnen Banden häufig unterscheiden. Eigenartigerweise ist dies bei der RSBH<sup>+</sup> jedoch kaum der Fall, was eine Besonderheit dieses Moleküls ist. Ein Vergleich der FTIR- und RR-Spektren in der Abbildung 1.11 demonstriert die Ähnlichkeit der beiden Spektren hinsichtlich der Intensität der Chromophorbanden. Gleichzeitig zeigt der Vergleich, dass im FTIR-Spektrum zusätzliche Banden auftauchen, die dem im RR-Spektrum ausgeblendeten Protein zuzuordnen sind.

Die eingehende Analyse solcher Schwingungsspektren ist eine hochkomplizierte Aufgabe, da die Schwingungsspektren eines Moleküls nicht nur von seiner Geometrie sondern auch von dem intramolekularen Kraftfeld beeinflusst werden. Insbesondere werden die Positionen und Intensitäten der einzelnen Banden in einem Schwingungsspektrum durch die Kopplung der diversen Schwingungen innerhalb einer moleku-



laren Gruppe und durch elektrostatische Wechselwirkungen mit ihrer Umgebung beeinflusst. Um die experimentell gewonnenen Schwingungsspektren interpretieren zu können, müssen sie einer umfassenden theoretischen Analyse unterzogen werden.

## 1.3 Theoretische Methoden und Beschreibungen

Aus einer theoretischen Analyse von Schwingungsspektren können im Prinzip Rückschlüsse auf die sterischen und elektrostatischen Wechselwirkungen sowie die Geometrie der beteiligten Molekülgruppen gezogen werden. Die Techniken zur Dekodierung dieser Daten sind jedoch noch nicht hinreichend weit entwickelt, um immer verlässliche Aussagen zu liefern. Da das vorrangige Ziel der vorliegenden Arbeit ist, Schwingungsspektren von Biomolekülen zu analysieren, werde ich deshalb, Grundlagen und einzelne Techniken zum Dekodieren von Schwingungsspektren etwas näher erläutern.

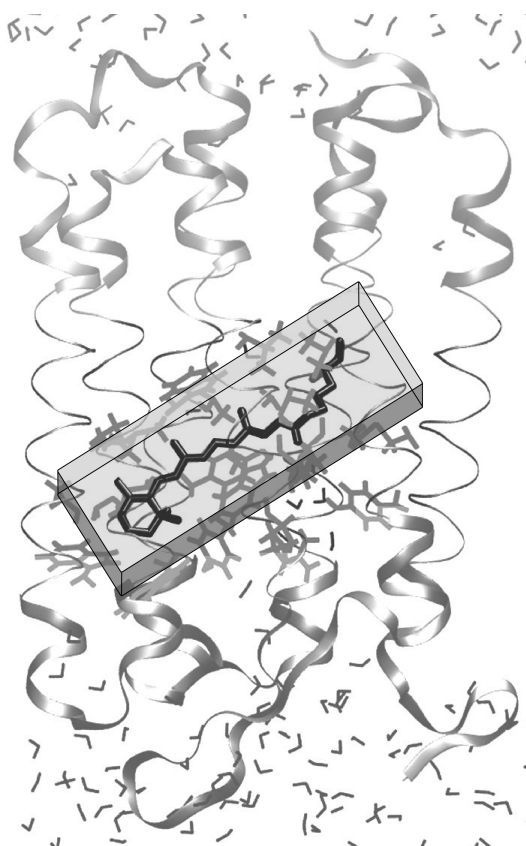
### 1.3.1 Methoden zur Berechnung von IR Spektren

Die Schwingungsspektren eines Moleküls werden durch sein intramolekulares Kraftfeld bestimmt, welches durch die Lösung der stationären elektronischen Schrödingergleichung zugänglich ist. In harmonischer Näherung muss dazu die Hessesche Matrix der zweiten Ableitungen der Grundzustandsenergie nach den Kernkoordinaten bei einer Gleichgewichtskonfiguration des Moleküls bestimmt werden [82]. Also erfordert die Berechnung der Schwingungsspektren von Molekülen zunächst ein geeignetes, d.h. hinreichend genaues quantenmechanisches Verfahren zur Berechnung des elektronischen Grundzustandes eines Moleküls. Durch die Entwicklung der Dichtefunktionaltheorie (DFT) [83, 84] erlaubt es der derzeitige Stand der Technik, die IR Spektren von isolierten Molekülen mit hoher Genauigkeit zu berechnen [85].

Ist ein Molekül stark polarisierbar und eingebettet in eine Umgebung aus polaren oder geladenen Molekülen, so muss der Einfluss der Umgebungselektrostatik auf die elektronische Wellenfunktion des Moleküls geeignet berücksichtigt werden [57], da äußere elektrische Felder das intramolekulare Kraftfeld eines polarisierbaren Moleküls beträchtlich modifizieren können. Für die Berechnung von IR Spektren von Molekülen in kondensierter Phase, z.B. von Molekülen in Lösung oder in einer komplexen Proteinumgebung, sind DFT Verfahren aufgrund ihres großen Rechenaufwands jedoch nicht geeignet. Folglich muss man auf die vereinfachte MM Modelle zurückgreifen, die annähernd elektrostatische Effekte der Umgebung berücksichtigen.

Eine entsprechende DFT/MM-Hybridmethode [56], welche die hochgenauen DFT-Rechnungen mit der Effizienz von molekülmechanischen Kraftfeldern kombiniert, wurde von M. Eichinger im Rahmen seiner Dissertation entwickelt und technisch als Kombination des MM-Programms EGO-MMII [86] mit dem DFT-Programm CPMD [87] realisiert. Dabei wird nur ein Teil des betrachteten Systems, mit einem vom CPMD-Programm zur Verfügung gestellten DFT-Verfahren genau beschrieben (DFT-Fragment) und das umgebende Lösungsmittel wird molekülmechanisch modelliert (MM-Fragment).

Für das BR zeigt Abbildung 1.12 eine mögliche Einteilung in ein DFT-Fragment (schwarz), welches den Chromophor und Lys216 umfasst, und ein MM-Fragment (grau), welches den Rest von BR überdeckt. Da das Programm CPMD die Elektronendichte des DFT-Fragments in eine Basis von ebenen Wellen entwickelt, die in einer sog. DFT-Box lokalisiert sind, wurde diese Box in die Abbildung 1.12 zusätzlich eingezeichnet. Im von mir verwendeten DFT/MM-Verfahren werden die kova-



**Abbildung 1.12:** Einteilung des BR in ein DFT- (schwarz) und ein MM-Fragment (grau). Zusätzlich ist die DFT-Box eingezeichnet, in der die Kohn-Scham-Orbitale nach ebenen Wellen entwickelt werden.

lenten Wechselwirkungen zwischen dem Protein und dem Chromophor durch die SPLAM-Methode (*Scaled-Position-Link-Atom-Methode*) erfasst. Hierbei wird das MM-Fragment durch eine geeignete Positionierung eines Wasserstoffatoms (Link-Atoms) abgesättigt. Auch für die Berücksichtigung der übrigen nicht-kovalenten Wechselwirkungen (van der Waals, Elektrostatik) wurden geeignete Verfahren entwickelt. Insbesondere wird dabei das vom MM-Fragment erzeugte elektrostatische Potential in den DFT-Hamiltonoperator importiert und so der Polarisierung des DFT-Fragments durch die MM-Umgebung Rechnung getragen. Umgekehrt wird die elektrostatische Rückwirkung des DFT-Fragments auf die MM-Umgebung im Rahmen der Partialladungsnaheung behandelt.

Mittlerweile wurde die angesprochene DFT/MM-Hybridmethode zur Berechnung der Schwingungsspektren von Molekülen in kondensierter Phase auf eine Reihe wichtiger Beispiele angewandt und es wurden die dabei einzusetzenden Rechenprotokolle genauer analysiert. Diese Anwendungen umfassen die Schwingungsspektren der Chinon-Kofaktoren  $Q_A$  und  $Q_B$  im photosynthetischen Reaktionszentrum von *Rb. Sphaeroides* [88] sowie von p-Benzochinon [89], Phosphationen [90] und Formaldehyd in Wasser [91, 92]. Insbesondere die letzteren Arbeiten haben eine Reihe wichtiger methodischer Fragestellungen, die durch den Versuch aufgeworfen werden, Schwingungsspektren von Molekülen in kondensierter Phase zu berechnen, grundlegend geklärt, während die übrigen Arbeiten vornehmlich anhand des Vergleichs mit experimentellen Daten und früheren Rechnungen die hohe Qualität der erzielbaren Resultate nachgewiesen haben.

Die inzwischen gut etablierten Protokolle [92] wie die instantane Normalmodenanalyse (INMA), die Fourier Transformation der Zeitkorrelationsfunktion (FTTCF) des Dipolmoments und das zeitaufgelöste, fehlerbereinigte generalisiertes Virial (TF-GV) erlauben eine hochpräzise Berechnung der Frequenzlagen, inhomogen verbreiteter Linienform sowie Intensitäten der IR-Banden im mittleren infraroten Spektralbereich. Dabei erweist sich die INMA-Methode als wesentlich billiger als FTTCF oder TF-GV, da sie lediglich den Einsatz von MM-MD-Simulationen erfordert, um ein thermisches Ensemble von Molekülstrukturen in der Lösung zu generieren. Für jede Struktur dieses Ensembles erfolgt anschließend die DFT/MM-Geometrieoptimierung und die Bestimmung der Hesse-Matrix in der festgehaltenen MM-Lösungsmittelumgebung. Durch Wiederholung dieser Prozedur für verschiedene Schnappschüsse aus einer MM-MD-Trajektorie lassen sich zeitabhängige Spektren berechnen. Ein weiterer Vorteil der INMA-Methode besteht darin, dass man durch die Normalmoden eine Beschreibung der Molekülbewegungen erhält. Außerdem lassen sich Isotopeneffekte besonders leicht durch Ändern der betroffenen Atommassen bei der Berechnung der massengewichteten Hesse-Matrix berücksichtigen.

Im Gegensatz basieren FTTCF und TF-GV auf sehr teuren DFT/MM-MD Simulationen von wenigstens einigen zehn Picosekunden. Die Vereinfachungen bei INMA haben zwar einen entscheidenden Vorteil hinsichtlich der Kosten und des Aufwandes bei der Berechnung von Schwingungsspektren, führen jedoch dazu, dass die Methode aufgrund der Vernachlässigung von Effekten des "motional narrowing" Linienbreiten um etwa 40 % [92] überschätzt.

Ferner erfordert INMA eine stabile MM-MD-Trajektorie. Im Falle des BR erwies sich dies als eine große Schwierigkeit, da die Kristallstruktur von BR im Laufe der in verschiedenen Arbeitsgruppen durchgeführten MD-Simulationen kollabierte [56, 93]. Wie gezeigt werden wird [1], beruht diese Instabilität auf der Vernachlässigung der Polarisierungseffekten durch konventionelle Kraftfelder. Im Folgenden möchte ich deswegen auf die theoretischen Grundlagen der Kraftfeldmethoden eingehen.

### 1.3.2 MM-MD-Simulationen

In MM-MD Simulationen werden die Atome der untersuchten Moleküle durch Massenspunkte repräsentiert und die Wechselwirkungskräfte aus einer analytisch vorgegebenen, geeignet parametrisierten Potentialfunktion  $E_{MM}(R)$ , dem sogenannten MM-Kraftfeld, berechnet. Zur Berechnung einer Trajektorie von Systemkonfigurationen  $\mathbf{R}$ , die vom Simulationssystem bei seiner thermischen Bewegung durchlaufen werden, löst man numerisch die Newtonschen Bewegungsgleichungen

$$m_i \frac{d^2}{dt^2} \mathbf{R}_i = -\nabla_i E_{MM}(\mathbf{R}), \quad i = 1, \dots, N \quad (1.1)$$

für alle Atome  $i$ .

Das Kraftfeld wird in üblichen MM-Modellen wie etwa OPLS [94], CHARMM [95], AMBER [96] gemäß

$$E_{MM}(\mathbf{R}) = E_{bonded}(\mathbf{R}) + E_{nonbonded}(\mathbf{R}) \quad (1.2)$$

in einen so genannten *bindenden Anteil*  $E_{bonded}(\mathbf{R})$ , der die lokalen Kräfte aus kovalenten Bindungen erfassen soll, und einen *nichtbindenden Anteil*  $E_{nonbonded}(\mathbf{R})$ , der die Wechselwirkungen chemisch nicht unmittelbar aneinander gebundener Atome beschreiben soll, aufgeteilt.

Nach dem CHARMM-Konzept werden die bindenden Wechselwirkungen  $E_{bonded}$  gemäß

$$E_{bonded}(\mathbf{R}) = E_{stretch}(\boldsymbol{\lambda}) + E_{angle}(\boldsymbol{\varphi}) + E_{dihed}(\boldsymbol{\theta}) \quad (1.3)$$

in Beiträge  $E_{stretch}(\boldsymbol{\lambda})$  für die Änderungen von Bindungslängen  $\boldsymbol{\lambda}$ ,  $E_{angle}(\boldsymbol{\varphi})$  für die Deformationen von Bindungswinkeln  $\boldsymbol{\varphi}$  und  $E_{dihed}(\boldsymbol{\theta})$  für die Beschreibung von Torsionswinkeländerungen  $\boldsymbol{\theta}$  aufgeteilt. alle Anteile der bindenden Wechselwirkungen durch einfache Potentialfunktionen beschrieben, und anschließend aufsummiert. Nach diesem Ansatz wird also angenommen, dass zwischen den internen Koordinaten  $\lambda$ ,  $\varphi$ , und  $\theta$  im jeweiligen Molekül keine Kopplungen existieren.

Unter Annahme kleiner Auslenkungen, die bei Zimmertemperatur für kovalente Bindungen in guter Näherung gegeben sind, kann der Bindungsterm  $E_{stretch}(\boldsymbol{\lambda})$  einfach durch eine Summe von harmonischen Potentialen approximiert werden:

$$E_{stretch}(\boldsymbol{\lambda}) = \sum_{i=1}^u \alpha_i (\lambda_i - \lambda_i^0)^2. \quad (1.4)$$

Dabei ist  $\lambda_i$  der aktuelle Abstand der zwei Atome,  $\lambda_i^0$  die Länge der Bindung in Ruhelage und  $\alpha_i$  die zugehörige Kraftkonstante.

Für den Bindungswinkelterm  $E_{angle}(\boldsymbol{\varphi})$  wird in CHARMM ebenfalls ein harmonischer Ansatz

$$E_{angle}(\boldsymbol{\varphi}) = \sum_{i=1}^v \beta_i (\varphi_i - \varphi_i^0)^2 \quad (1.5)$$

gewählt, der vermittelt der Kraftkonstanten  $\beta_i$  die Zunahme der Molekülenergie bei Auslenkungen der Bindungswinkel aus ihren jeweiligen Gleichgewichtswerten  $\varphi_i^0$  beschreibt.

Schließlich wird der Torsionsterm  $E_{dihed}(\boldsymbol{\theta})$  unter Ausnutzung der Periodizität eines Torsionswinkels  $\theta$  durch eine Kosinus-Entwicklung

$$E_{dihed}(\boldsymbol{\theta}) = \sum_{i=1}^w \sum_{m=1}^6 \gamma_{i,m} [1 + \cos(m\theta_i + \delta_{i,m})] \quad (1.6)$$

dargestellt.

Der Definition der nichtbindenden Wechselwirkungen aus Glg. (1.2) liegt die Annahme zu Grunde, dass alle Wechselwirkungen zwischen direkt (und indirekt über ein Zwischenatom) chemisch gebundenen Atomen schon durch die Modellpotentiale  $E(\boldsymbol{\lambda})$  und  $E(\boldsymbol{\varphi})$  für die Bindungslängen und -winkeländerungen erfasst sind. Daher müssen zusätzlich noch all jene Wechselwirkungen zwischen den Atomen eines Moleküls beschrieben werden, die weder durch  $E(\boldsymbol{\lambda})$  noch durch  $E(\boldsymbol{\varphi})$  noch durch die 4-Atomwechselwirkungen  $E(\boldsymbol{\theta})$  erfasst sind.

Die fehlenden Wechselwirkungen können nun durch eine Schaltfunktion

$$\epsilon(i, j) = \begin{cases} 0, & \text{falls Atom } i \text{ von Atom } j \text{ höchstens um 2 Bindungen entfernt ist,} \\ 1 & \text{sonst,} \end{cases}$$

angezeigt werden, die immer dann den Wert Eins annimmt, wenn fehlende Energieteile zu berücksichtigen sind.

Wie es bei MM-Kraftfeldern üblich ist, wird angenommen, dass sich diese gemäß

$$E_{\text{nonbonded}}(\mathbf{R}) = E_{\text{vdW}}(\mathbf{R}) + E_{\text{Coulomb}}(\mathbf{R}) \quad (1.7)$$

aus Van der Waals (VDW) und elektrostatischen Beiträgen zusammensetzen. Dabei wird die VDW-Wechselwirkung im CHARMM-Kraftfeld durch ein 12-6 Lennard-Jones-Potential

$$E^{\text{vdW}}(\mathbf{R}) = \sum_{i < j} \epsilon(i, j) \left[ \frac{C_{i,j}^{12}}{|\mathbf{R}_i - \mathbf{R}_j|^{12}} - \frac{C_{i,j}^6}{|\mathbf{R}_i - \mathbf{R}_j|^6} \right] \quad (1.8)$$

beschrieben, in dem der  $\mathbf{R}^{-12}$ -abhängige Teil die Pauli-Abstoßung sich überlappenden abgeschlossener Elektronenschalen und der mit einem  $\mathbf{R}^{-6}$ -Verhalten langreichweitigere Anteil die attraktive Dispersionswechselwirkung modelliert. Zur Erfassung der elektrostatischen Wechselwirkungen werden im CHARMM-Kraftfeld den Atomen partielle Ladungen  $q_i$  zugewiesen, welche beispielsweise die Polarisierung chemischer Bindungen zwischen Atomen unterschiedlicher Elektronegativität modellieren. Bei Verwendung atomarer Einheiten wirkt zwischen zwei solchen Ladungen die Coulomb-Wechselwirkung

$$E^{\text{Coulomb}}(\mathbf{R}) = \sum_{i < j} \epsilon(i, j) \frac{q_i q_j}{4\pi\epsilon\epsilon_0 |\mathbf{R}_i - \mathbf{R}_j|} . \quad (1.9)$$

Wie man sieht, beruht ein solches MM-Kraftfeld also auf sehr einfachen Modellvorstellungen, welche die tatsächlichen physikalischen Gegebenheiten nur näherungsweise widerspiegeln können. Die praktische Anwendung der MM-Methode setzt nun voraus, dass die Vielzahl von Parametern, die in die Gleichungen (1.4)–(1.9) eingehen und die das CHARMM Kraftfeld endgültig spezifizieren, bekannt sind.

Zur Festlegung der Parameter werden in der Regel quantenmechanische Rechnungen an Modellsubstanzen, gelegentlich aber auch experimentelle Daten verwendet. Um die Parameterzahl zu begrenzen und wenigstens einigermaßen überschaubar zu halten, werden bei CHARMM ferner so genannte *Atomtypen* definiert, die durch Atome eines gegebenen chemischen Elements in einer gegebenen Bindungsumgebung und

in einem gegebenen Hybridisierungszustand definiert sind. Das CHARMM-Kraftfeld umfasst Parameter für etwa 50 verschiedene Atomtypen, die zu den wenigen in Proteinen vorkommenden chemischen Elementen gehören. Für jeden Atomtyp liefert es eine Beschreibung der bindenden Wechselwirkungen im angenommenen Bindungsmotiv durch Vorgabe der Gleichgewichtswerte und Kraftkonstanten für Bindungslängen, -winkel und -torsionen. Auch die nichtbindenden VDW Wechselwirkungen werden mit Parametern für die diversen Atomtypen spezifiziert. Da die Ladungsverteilung in einem Molekül jedoch nicht einfach durch lokale Effekte erklärbar ist, werden die für die Beschreibung der Elektrostatik benötigten Partialladungen dagegen unabhängig von den Atomtypen gewählt. Sie werden vielmehr für jedes Atom eines Moleküls oder einer molekularen Teilgruppe eines Makromoleküls durch quantenmechanische Berechnung festgelegt.

In nichtpolarisierbaren Kraftfeldern, wie CHARMM, werden *fixierte* Partialladungen eingesetzt, die im besten Falle, wie etwa beim Wassermolekül, versuchen, Polarisierungseffekte im Rahmen der Molekularfeldnäherung zu berücksichtigen. In der Regel werden diese Effekte jedoch gänzlich vernachlässigt und es werden die Partialladungen aus Rechnungen an isolierten Molekülen bestimmt. Für eine Diskussion der Bedeutung von Polarisierungseffekten in Proteinen und eine Abschätzung der Fehler, welche aus ihrer Vernachlässigung resultieren, möchte ich auf den Beitrag von Tavan et al. [97] verweisen.

Es sei an dieser Stelle angemerkt, dass wegen der außerordentlich großen Polarisierbarkeit des Retinalchromophors, die Ladungsverteilung und die Werte der Kraftkonstanten in diesem Molekül stark von der Umgebungselektrostatik abhängen [54]. Aufgrund der großen Empfindlichkeit der RSBH<sup>+</sup> gegenüber Änderungen der Umgebungselektrostatik kann daher erwartet werden, dass bei einer DFT/MM Hybridberechnung (vgl. Abb. 1.12) ihrer Schwingungsspektren auch kleinste Änderungen der MM-Elektrostatikmodellierung in der Chromophorumgebung zu deutlichen Effekten in den berechneten Spektren führen werden. Diese Spektren können daher als Sonden für eine Beurteilung der MM-Beschreibung der Umgebungselektrostatik dienen und sollten so auch über Fehler im MM-Kraftfeld Auskunft geben können.

## 1.4 Ziele und Gliederung dieser Arbeit

Mit der vorliegenden Arbeit sollen zunächst unter Einsatz der INMA-Methode, Daten aus einer Vielzahl schwingungsspektroskopischer Untersuchungen am Retinalchromophor des BR [20, 27–29] einer genaueren theoretischen Erklärung zugeführt werden. Zwar wurden auch früher schon beträchtliche Anstrengungen unternommen,

die Schwingungsspektren des BR-Chromophors zu analysieren [98–101], jedoch haben die Komplexität des BR, die starke Polarisierbarkeit seines Retinalchromophors [38, 98] und die entsprechend komplizierten Protein-Chromophor-Wechselwirkungen, sowie die unzureichende Genauigkeit der früher eingesetzten theoretischen Rechenverfahren [38] bis heute eine hinreichend genaue Beschreibung dieser Spektren verhindert. Die vorliegende Arbeit soll nun zur Beantwortung der Frage beitragen, mit welcher Qualität die IR-Spektren des Retinals *in situ* bei seiner sorgfältigsten Modellierung durch Anwendung der INMA-Methode berechnet werden können.

Wie bereits aus den bisherigen Diskussionen in dieser Einleitung folgt, erfordert eine genauere Berechnung der IR Spektren mittels INMA eine stabile MD-Trajektorie. Aus einer Reihe der Vorarbeiten zur BR [56, 93, 100, 101] ist jedoch bekannt, dass die Kristallstruktur des BR während der MD-Simulationen mit den standardmäßig eingesetzten Kraftfeldern nicht stabil bleibt. Die Klärung der Ursache für diese Instabilität war das nächste Ziel meiner Arbeit. Dazu habe ich die CHARMM-Parametrisierung der Elektrostatik in der Chromophor-Bindungstasche des BR geeignet, d.h. unter Einbeziehung der Polarisierbarkeit, modifiziert und die resultierenden Effekte analysiert.

Die darauf folgende Fragestellung war die Aufklärung der bereits erwähnten Unterschiede zwischen den NMR- und Röntgenbeugungsstrukturen des BR. Um diese Frage zu beantworten, wurde von meinem Kollegen Robert Denschlag eine Variante der Replica Exchange Methode [102, 103] entwickelt [1], die es gestattet den Konformationsraum kleiner Gruppen effektiv abtasten und damit Einblicke in die Wahrscheinlichkeitsverteilungen verschiedener Konformationen liefern kann.

Für die methodische Entwicklung der Berechnung der IR-Spektren des Chromophors *in situ* wurde zunächst geklärt, welchen Einfluss die Parametrisierung des MM-Fragments, insbesondere seiner Elektrostatik, auf die berechneten IR-Spektren hat. Schließlich wurde geklärt, wie stark Fluktuationen des elektrischen Feldes, die durch thermische Bewegungen der BR-Atome verursacht werden, die Normalmoden und zugehörigen Frequenzen des Chromophors beeinflussen können. Insgesamt wurde durch Vergleich mit den experimentellen Spektren untersucht, mit welcher Qualität diese Spektren bei sorgfältigster Modellierung reproduziert werden können. Die entsprechenden Ergebnisse dieses Teils meiner Arbeit habe ich zusammen mit Robert Denschlag, Gerald Mathias und Paul Tavan in den Artikeln [1] und [2] ausführlich beschrieben. Diese beiden Artikel wurden zur Veröffentlichung angenommen und sind als Abschnitt 2.1 bzw. 2.2 in die vorliegende Arbeit eingebunden.

Ein weiteres Ziel meiner Arbeit war, Daten aus den zeitaufgelösten Ultrakurzzeit-Spektroskopie-Experimenten am  $\beta$ -Hairpinpeptid mit theoretischen Daten zu ver-



gleichen und die Entfaltungsdynamik des  $\beta$ -Hairpinpeptides nach der Anregung des Photoschalters aufzuklären. Um diese Spektren zu interpretieren, habe ich mittels der DFT/MM-Methode die AmidI-Banden von zwei Faltungsintermediaten berechnet und mit experimentell beobachteten Spektren verglichen. Die entsprechenden Resultate habe ich zusammen mit Paul Tavan, Robert Denschlag, den Kollegen von der Arbeitsgruppe W. Zinth und von der Arbeitsgruppe L. Moroder in dem Artikel [3] publiziert, der als Kapitel 3 in die vorliegende Arbeit eingebunden ist.

Das abschließenden Kapitel 4 enthält eine ausführliche Zusammenfassung der Arbeit und skizziert offene Fragen.



## 2 Berechnung der IR-Spektren des Retinalchromophors von BR

### 2.1 Stabilisierung der Struktur des Chromophorbindungstasche in MD-Simulationen durch Berücksichtigung von Polarisisationseffekten

Um DFT/MM Berechnungen der Infrarotspektren des BR-Chromophors zu ermöglichen, ist ein stabiles MM-MD Simulationsmodell des Chromophorbindungstasche nötig. Im nachfolgend abgedruckten Manuskript

Galina Babitzki, Robert Denschlag, and Paul Tavan : „Polarization effects stabilize bacteriorhodopsin's chromophore binding pocket: A molecular dynamics study“,

das im *Journal of Physical Chemistry* publiziert ist, wird dieses Problem analysiert. Es wird zunächst gezeigt, dass die Chromophorbindungstasche bei MM-MD Simulationen mit den gängigen nichtpolarisierbaren Kraftfeldern kollabiert. Durch DFT/MM Rechnungen wird ein Satz von strukturadaptierten Partialladungen für die Residuen der Bindungstasche abgeleitet, welche die Effekte der dort stattfindenden wechselseitigen Polarisierung erfassen. Dieses neue polarisierte Kraftfeld erhält, wie gezeigt wird, die Struktur der Bindungstasche und schafft so die Voraussetzung für die Berechnung von Spektren.

Dank des Einsatzes meines Kollegen Robert Denschlag, konnte zusätzlich ein neuartiges Hamiltonsches Replica Austauschverfahren entworfen werden, mit dem der Nachweis gelang, dass die Lysinkette, über welche der Chromophor kovalent an das Apoprotein gebunden ist, bei Raumtemperatur mehrere metastabile Konformationen aufweist. Damit kann eine Vielzahl experimenteller Daten zur konformationellen Heterogenität von BR erklärt werden.

Zum Manuskript gehört Zusatzmaterial in Form ergänzender Tabellen, Graphiken und Texte, die im Anschluss an das Manuskript ebenfalls abgedruckt sind.

# Polarization Effects Stabilize Bacteriorhodopsin's Chromophore Binding Pocket: A Molecular Dynamics Study

G. Babitzki, R. Denschlag, and P. Tavan\*

Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität, Oettingenstr. 67, 80538 München, Germany

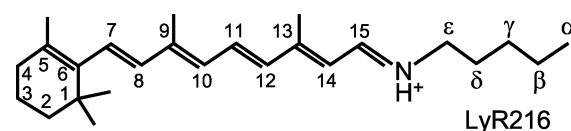
Received: March 18, 2009; Revised Manuscript Received: May 25, 2009

Hybrid methods, which combine a quantum mechanical description of a chromophore by density functional theory (DFT) with a molecular mechanics (MM) model of the surrounding protein binding pocket, can enable highly accurate computations of the chromophore's *in situ* vibrational spectra. As a prerequisite, one needs a MM model of the chromophore–protein complex, which allows a correct sampling of its room-temperature equilibrium fluctuations by molecular dynamics (MD) simulation. Here, we show for the case of bacteriorhodopsin (BR) that MM–MD descriptions with standard nonpolarizable force fields entail a collapse of the chromophore binding pocket. As demonstrated by us, this collapse can be avoided by employing a polarized MM force field derived by DFT/MM hybrid computations. The corresponding MD simulations, which are complemented by a novel Hamiltonian replica exchange approach, then reveal a structural heterogeneity within the binding pocket of the retinal chromophore, which mainly pertains to the structure of the lysine chain covalently connecting the retinal chromophore with the protein backbone.

## Introduction

The well-known membrane protein bacteriorhodopsin (BR) acts as a light-driven proton pump in the cell membrane of the archaeobacterium *Halobacterium salinarum*. Since its detection by Oesterhelt and Stoekenius,<sup>1,2</sup> an enormous amount of scientific effort has been devoted to uncover the physical mechanisms governing the structure and functional dynamics of this small molecular machine (for reviews see, e.g., refs 3 and 4). The popularity of BR in biophysics and beyond is due to its exceptional stability, its simple accessibility, and the rapid turnover of its light-triggered proton pump cycle: A few milliseconds after light absorption, the proton pumping is complete. The pumping proceeds through several intermediate states, whose optical and vibrational spectra are markedly different and show lifetimes ranging from pico- up to milliseconds. This spectroscopically detectable so-called photocycle is kinetically coupled to a vectorial proton transport across the membrane. The transport involves a series of de- and reprotonations of the chromophore and of various aspartic acids. As a consequence of BR's favorable properties, many spectroscopic methods have first been tested and established using BR as a sample system.

Particularly for the demonstration of how one can use the various variants of vibrational spectroscopy for a detailed characterization of dynamical processes in biomolecules, BR has been (and still is; see, e.g., refs 5–7) a testing ground.<sup>8</sup> Here, resonance Raman spectroscopy can shed light on structural transformations of the chromophore,<sup>9–11</sup> whereas infrared (IR) difference techniques additionally enable a detailed monitoring of light-induced changes within the surrounding protein matrix.<sup>12</sup> Both techniques can be applied at a very high (up to fs) time resolution to chromoproteins at ambient temperatures and therefore enable the observation of dynamical processes under quasi-native conditions.



**Figure 1.** Chemical structure of the BR chromophore: In the light-adapted state BR<sub>568</sub>, the chromophore is the protonated Schiff base of all-*trans* retinal (RSBH<sup>+</sup>), which is covalently linked to the protein through Lys216, thus constituting a special residue called LyR216.

Unfortunately, it is generally not easy to decode the wealth of structural information contained in the vibrational spectra. Exceptions are, e.g., the spectrally well-separated bands of protonated carboxylic acids, whose involvement in proton transfer processes can therefore be observed in a kinetically resolved fashion.<sup>13,14</sup> In contrast, a particularly hard case is the BR chromophore when it comes to determine the structural meaning of the associated vibrational bands as they are observed in the course of the photocycle.<sup>15–19</sup> Nevertheless, for the light-adapted initial state BR<sub>568</sub> of this cycle, the chromophore structure is quite precisely known<sup>9</sup> and is the protonated Schiff base of all-*trans* retinal depicted in Figure 1.

The positive charge, which is introduced into the retinal molecule by the protonation, delocalizes into the conjugated  $\pi$ -electron system and this mobile charge is the cause for the huge polarizability of an RSBH<sup>+</sup>. Because of this polarizability, local electric fields generated by a condensed-phase environment (like, e.g., by a chromophore binding pocket of a protein) can steer the properties of an embedded RSBH<sup>+</sup> and, here, particularly the bonding properties within the  $\pi$ -conjugated chain. The optical and IR spectra of the chromophore are shaped by these properties. Because structural changes of the environment generally lead to changes of the local electric fields, such changes shift the spectral positions of prominent absorption bands in the optical and IR spectral regions.<sup>19,20</sup> The extraordinary sensitivity, with which an RSBH<sup>+</sup> changes its electronic structure in response to external fields, is the reason why this chromophore is an excellent spectroscopic probe for dynamical

\* Corresponding author. E-mail: tavan@physik.uni-muenchen.de. Phone: +49-89-2180-9220. Fax: +49-89-2180-9202.

processes in a protein. On the other hand, this sensitivity is also the main cause for the challenge, which this chromophore poses to theoretical descriptions.

**Computational Approaches to the Chromophore's IR Spectra.** About 15 years ago, density functional theory<sup>21,22</sup> (DFT) was converted into an easily accessible tool for computing the ground-state properties of isolated molecules.<sup>23</sup> Since then, accurate computations of gas-phase vibrational spectra have become standard software applications.<sup>24–26</sup> However, for molecules embedded in polar solvents or in proteins, the strong electrostatic interactions with the environment have to be included in the DFT Hamiltonian.<sup>27</sup> Because the electrostatic interactions are long-ranged and because the polarizability of an RSBH<sup>+</sup> is very large, the condensed-phase environments capable of polarizing such a chromophore comprise too many atoms to be included in a DFT treatment at a manageable computational cost. Therefore, one has to resort to simplified molecular mechanics (MM) models that can approximately account for the electrostatics of the chromophore's environment.

A corresponding DFT/MM hybrid method was designed by Eichinger et al.<sup>28</sup> to enable accurate computations of a chromophore's vibrational spectra in a condensed-phase environment. Subsequent applications of this method to quinones in the reaction center of *Rb. sphaeroides* and in aqueous solution,<sup>29,30</sup> as well as to various molecules in solution,<sup>31–36</sup> demonstrated the merits and remaining deficiencies of this hybrid method. Like all other hybrid methods, which describe a small fragment of a simulation system by a quantum mechanical (QM) Hamiltonian and the remainder by an MM force field (see ref 28 for extended references), also the quoted DFT/MM approach conceptually derives from the seminal work of Warshel and Levitt.<sup>37</sup> Its key merit is a careful treatment of covalent bonds linking the QM and MM fragments which removes corresponding artifacts from the vibrational spectra of the QM fragment, hampering previous methods (see ref 28 for details).

As discussed in a recent review,<sup>27</sup> the mid-IR spectra of molecules in a condensed-phase environment can be calculated by DFT/MM at a comparable accuracy as the gas-phase spectra by DFT, as long as these molecules are at most mildly polar. In contrast, charged solute molecules like, e.g., phosphate ions<sup>31,38</sup> will strongly polarize their solvent environment and will thus be exposed to a strong reaction field that can cause very large shifts of the solute's vibrational bands. In conventional MM force fields, however, the dynamic electronic polarizability is neglected. Thus, for charged solute molecules, such force fields underestimate the strength of the reaction field and the associated solvatochromic shifts of the vibrational bands.<sup>27</sup>

Using meanwhile established computational protocols,<sup>33</sup> one can accurately calculate by DFT/MM not only the frequencies but also the inhomogeneously broadened shapes and intensities of the mid-IR bands of a solute molecule. Among these protocols, the so-called "instantaneous normal-mode analysis" (INMA) is computationally the least demanding. Here, one employs a standard MM molecular dynamics (MD) simulation to generate a set of solvent–solute structures, which should be representative of the equilibrium fluctuations at room temperature. Subsequently, the solvent cages of these MM–MD snapshots are kept fixed and a series of vibrational line spectra is calculated for the solute by means of DFT/MM normal-mode analysis. Other protocols, in contrast, need a very costly DFT/MM–MD simulation that has to cover at least several tens of picoseconds (see ref 33 for a comparative discussion). Thus, the condensed-phase vibrational spectra of large molecules (like

a peptide in solution<sup>34</sup> or the RSBH<sup>+</sup> in BR) can be currently computed only through the INMA protocol in an accurate and computationally manageable fashion.

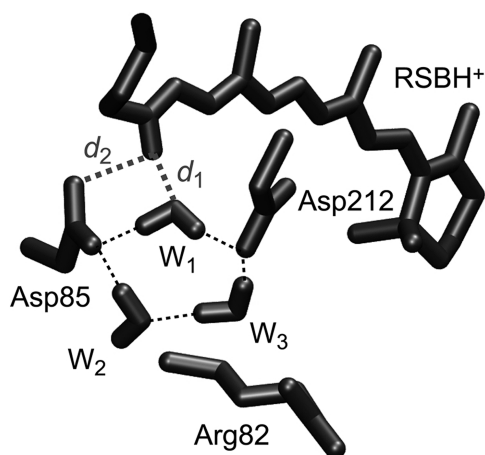
One may ask why accurate computations of the vibrational spectra of the BR chromophore in the various states of the pump cycle are still of interest (and even required, if one wants to gain a physical understanding of the pump mechanism in BR). This question will be addressed in part II of this work (DOI 10.1021/jp902432e), which presents our DFT/MM results on the IR spectra of the BR chromophore obtained by the INMA approach. Instead, here, we address the question as to how one can obtain by MM–MD simulations a set of structural snapshots for the RSBH<sup>+</sup> in the BR binding pocket, which correctly represents the room-temperature equilibrium ensemble and, thus, provides a valid starting point for part II (DOI 10.1021/jp902432e). In fact, this task turned out to be technically much harder than the subsequent DFT/MM computation of the IR spectra from a given snapshot ensemble.

This paper is organized as follows: First, we review previous MM–MD descriptions of BR<sub>568</sub> which, by applying standard force fields, tried to model the equilibrium conformational fluctuations. We show that these descriptions fall short of explaining the corresponding experimental evidence. By arguing that this failure is due to the neglect of the electronic polarizability in standard MM force fields, we offer our working hypothesis. Extended MD simulations of BR, which compare standard force fields with polarized force fields derived from DFT/MM calculations, subsequently serve us to provide evidence for the validity of our hypothesis. Here, we first describe in some detail the computational methods employed in the simulations, in the DFT/MM computations of polarized force fields, and in the statistical data analysis. Next, we present the results and particularly discuss the issue of structural heterogeneity emerging from the simulation data. A short summary and outlook concludes the paper.

## BR<sub>568</sub>—A Challenge for MM–MD

Already in 1999, the complexities of the task to generate by MM–MD a set of snapshot structures, which correctly represents the room-temperature structural ensemble within the chromophore binding pocket of BR<sub>568</sub>, started to show up in the thesis of Eichinger.<sup>39</sup> This thesis reports first attempts to describe the IR spectra of the BR chromophore by a DFT/MM approach. At that time, the high-resolution X-ray structure 1C3W of BR was still unknown.<sup>40</sup> According to this structure and as shown in Figure 2, in BR, the chromophore binding motif is composed of four key charged residues, which form a neutral but strongly quadrupolar charge distribution. This charge distribution is located near the center of the membrane. As follows from dynamics modelings (see, e.g., further below), the charge distribution is stabilized by a hydrogen bonded network. This network is generated by three water molecules, which are also shown in Figure 2 with their oxygen atoms at the crystallographic locations.

Instead of this high-resolution structure, Eichinger<sup>39</sup> had to choose lower resolution models like those of Grigorieff et al.<sup>41</sup> or Essen et al.<sup>42</sup> as starting points for the computation of IR spectra. Because these models contained no water molecules at all, such molecules had to be added by cavity filling. In this process, a water molecule was always initially modeled into the position of the water molecule W<sub>1</sub> in Figure 2,<sup>39</sup> because room-temperature spectroscopic evidence had clearly indicated that the Schiff base proton must be stably hydrogen bonded to a water molecule. This hydrogen bonding motif had been first



**Figure 2.** Hydrogen bonded network (black dashed lines) linking the key charged residues and the three water molecules in the chromophore binding pocket of BR. This network follows from the high-resolution (1.55 Å) crystal structure 1C3W<sup>40</sup> by applying the modeling and simulation procedures described in Methods. The distances (gray dashed lines) between the Schiff base proton and the oxygen atom of the closest water molecule ( $d_1$ ) or the closest oxygen of Asp85 ( $d_2$ ), respectively, can serve as measures for the stability of the depicted structure.

detected by Hildebrandt and Stockburger<sup>43</sup> using a clever argument on the changing line width of the C=N stretching band observed in the resonance Raman spectrum of BR<sub>568</sub> upon H<sub>2</sub>O/D<sub>2</sub>O exchange. The subsequent confirmation<sup>44</sup> by high-resolution, solid-state <sup>15</sup>N NMR has additionally shown that the hydrogen bond between the Schiff base proton and a water molecule stably persists on the extended time scale of such measurements.

In the study of Eichinger,<sup>39</sup> the apoprotein was modeled by the standard CHARMM22 force field<sup>45</sup> and the chromophore as well as part of the lysine chain (cf. Figure 1) by DFT. Encouragingly, the IR line spectra of the chromophore calculated by DFT/MM for the quoted structural models were much closer to the well-known experimental data<sup>46,47</sup> than IR spectra calculated by DFT for an isolated RSBH<sup>+</sup>. Nevertheless, the calculated spectra also showed marked differences (i) from each other and (ii) from the experimental spectra, indicating (i) that small differences among the chosen molecular models are magnified by the polarizability of the RSBH<sup>+</sup> into sizable spectral line shifts and (ii) that the chosen models are most probably not quite correct.

#### Is BR's Chromophore Binding Pocket Stable in MM–MD?

Line spectra derived by DFT/MM from a static model of a low-temperature protein structure are not the kind of theoretical predictions that one should compare to room-temperature vibrational spectra. Instead, one should try to get access to the room-temperature ensemble of chromoprotein configurations to additionally model the fluctuations of the vibrational lines and, thus, to get access to the band shapes. However, when Eichinger<sup>39</sup> tried to generate such an ensemble by MM–MD simulations for water soaked BR models,<sup>41,42</sup> he observed that these structures “collapsed”. Thus, an INMA computation of the IR spectra was not feasible and either the employed MM force field (CHARMM22), the positions of the added water molecules, the experimental starting structures, or all of that seemed to be too inaccurate for generating stable equilibrium ensembles of BR by MM–MD.

Meanwhile, the 1.55 Å structure of Luecke et al.<sup>40</sup> has removed the uncertainties concerning the experimental low-temperature structure of BR<sub>568</sub>. Furthermore, the NMR analysis of BR carried out by Patzelt et al.<sup>48</sup> has shown that the structural

arrangement of the four charged groups and three water molecules depicted in Figure 2 is stably maintained also at room temperature. Finally, the recent FTIR study of Garczarek and Gerwert<sup>5</sup> has clearly confirmed that in BR<sub>568</sub> this structural arrangement remains rock-solid at room temperature.

Interestingly, huge unconstrained MM–MD simulations of BR trimers embedded in a periodic membrane/water solvent model showed a different picture.<sup>49–51</sup> For instance, Kandt et al.<sup>50</sup> stated for “the triple water cluster... near the Schiff base” (and depicted in Figure 2) that the “MD simulation shows highly mobile waters with fluctuations of 4 Å per individual water molecule”. Furthermore, the pictures, which are given in the quoted paper as illustrations for the resulting “water density” (Figure 4a and b in ref 50), clearly indicate that the Schiff base proton was hydrogen bonded to Asp85 in this simulation. Thus, the hydrogen bond of the Schiff base group with a water molecule, which stably persists in BR on the time scale of NMR measurements, actually decayed within the 5 ns of the quoted MD simulation. Grudinin et al.<sup>51</sup> using a similarly extended simulation model likewise found that the water molecule denoted as W<sub>1</sub> in Figure 2 left its crystallographic position. Also, the MM–MD simulations of Eichinger<sup>39</sup> consistently showed a decay of the initially modeled hydrogen bond between the Schiff base proton and the water molecule W<sub>1</sub> followed by the formation of a new hydrogen bond between the Schiff base proton and Asp85. Eichinger<sup>39</sup> considered this conformational transition as a structural “collapse” and concluded that his MM–MD simulations cannot provide any insight into the equilibrium structure and fluctuations of the BR binding pocket. Kandt et al.<sup>50</sup> and Grudinin et al.,<sup>51</sup> while stating the difference of their MM–MD simulations with the X-ray data, attributed this difference solely to the simplicity of the employed water model and were apparently quite content with the fact that W<sub>1</sub> stayed at least near its crystallographic position during their 5 and 2.5 ns simulations, respectively.

Concerning the MM–MD methodology, it is important to note that Kandt et al.<sup>50</sup> employed the GROMACS<sup>52,53</sup> force field for the apoprotein and the associated simple point charge (SPC) model for the water molecules, that Grudinin et al.<sup>51</sup> used AMBER for the protein and the three-point transferable interaction potential<sup>54</sup> [TIP3P] for the water molecules, and that Eichinger<sup>39</sup> used CHARMM22 combined with TIP3P.<sup>45</sup> Despite these force field differences, the same type of structural decay occurred in all three studies.

That this type of structural decay seems to be a general feature of conventional nonpolarizable MM force fields can be concluded from the MD study of Hayashi et al.<sup>55</sup> who noted “that the TIP3P/AMBER force field, in which nonbonding interactions are described by only classical pairwise potential functions, cannot even reproduce the conformation of Wat402 observed in the X-ray ground-state structure; after the MD equilibration, the Schiff base N–H bond directly attaches to O<sub>2</sub> of Asp85.” With GROMACS, CHARMM22, and AMBER, the three most widespread conventional MM force fields for proteins thus showed the same failure in MD simulations of BR: The water molecule, which according to the available experimental evidence is stably attached to the Schiff base proton, apparently does not want to stay at this position. Instead, the RSBH<sup>+</sup> and Asp85 are predicted to form a stable ion pair. We will denote the resulting non-native conformation by C<sub>nn</sub>.

Therefore, one expects that this structural decay into conformation C<sub>nn</sub> will also occur in MM–MD simulations with the TIP3P/CHARMM22 force field, in which a BR model



deriving from the high-resolution structure 1C3W is employed.<sup>40</sup> As we will show further below, this is actually the case.

**Possible Sources of the Apparent Force Field Deficiencies.** Hayashi et al.<sup>55</sup> fixed the problem of structural decay by introducing additional local potential functions capable of constraining the water molecule  $W_1$  at the experimental position. Furthermore, they used instead of the static TIP3P water model the polarizable potential function of Rick et al.<sup>56</sup> These choices for fixing the apparent force field bugs were motivated by a certain interpretation of QM/MM results,<sup>57</sup> which attributed the considerable stability of the hydrogen bond between the RSBH<sup>+</sup> and the water molecule  $W_1$  mainly to local charge transfer effects. However, because the quoted QM/MM study<sup>57</sup> did not compare the native arrangement shown in Figure 2 with the ion pair structure (RSBH<sup>+</sup>...Asp85<sup>-</sup>), it cannot answer the question of why the conventional MM force fields erroneously prefer the ion pair over the experimentally observed hydrogen bonded network. Therefore, it is uncertain whether this MM–MD failure is caused by an inaccurate modeling of certain local effects (refs 55 and 57) or is due to more general properties of conventional MM force fields.

Because we generally hold the view that the main shortcoming of conventional MM force fields is their systematic neglect of the dynamic electronic polarizability,<sup>58,59</sup> in this paper, we will address the question as to whether the inclusion of polarization effects into the protein force field can stabilize the chromophore binding pocket of BR at the experimental structure in extended MM–MD simulations. In the given case, such a stabilization by polarization seems likely, because the key structural motif consists of four charged residues in the low-dielectric center of a membrane, where the associated Coulomb fields are poorly screened and, thus, are strong sources of electronic polarization.

To keep the focus on the polarization of the protein, we will maintain the nonpolarizable TIP3P model for the water molecules in our extended MD simulations. We will, however, check by DFT/MM calculations to what extent the electronic polarizability of the three water molecules in the chromophore binding pocket can contribute to the stabilization of the experimental structure. Note in this context that conventional water models like TIP3P include a mean field description of polarization effects through the assumption of a “condensed-phase” dipole moment that is much larger than the dipole moment of an isolated water molecule (see ref 35 for a recent discussion of water polarizability). Whether this mean polarization applying to bulk water suffices for the three water molecules in the hydrophobic center of BR is to be seen.

## Methods

**Setup of a Simulation Model for BR.** For the construction of a BR simulation model, we chose the 1.55 Å X-ray structure 1C3W as our starting point.<sup>40</sup> Trial coordinates for the hydrogen atoms and for residues, which are missing in the 1C3W data set (Thr157–Glu161, Glu232, Ala233), were added with the help of the MD program X-PLOR.<sup>60</sup> Next, the CHARMM22 force field<sup>45</sup> was assigned to the protein, the TIP3P model<sup>54</sup> to the crystallographic water molecules, and an MM model to the RSBH<sup>+</sup> (see further below).

With the exception of the four charged residues in the center of the membrane (see Figure 2) and of two glutamates, which are located in the extracellular proton release channel relatively close to the RSBH<sup>+</sup>, all other titratable groups were chosen as neutral. There are several titratable groups (including the C- and N-termini) near the two water-exposed surfaces of BR,

which should be charged at a pH of 7. The neutralization of these groups serves as a crude model for the strong dielectric shielding of their Coulomb potentials by the aqueous environment. Near the center of the membrane, there is no such shielding and, therefore, in this region the charges of titratable groups were chosen as given by experimental evidence.<sup>61–63</sup> Correspondingly, a negative charge was assigned to Asp85 and Asp212, a positive charge to Arg82 and LyR216, and no net charge to Asp96 and Asp115. To account (i) for the presence of a delocalized proton in the extracellular proton release channel of BR<sup>5,64</sup> and (ii) for the partial dielectric shielding in the associated region at the boundary between the hydrophobic core of the membrane and the water-exposed extracellular surface, a charge of  $-0.5 e$  was assigned to each of the two glutamates (Glu194 and Glu204), marking the exit of the proton release channel.

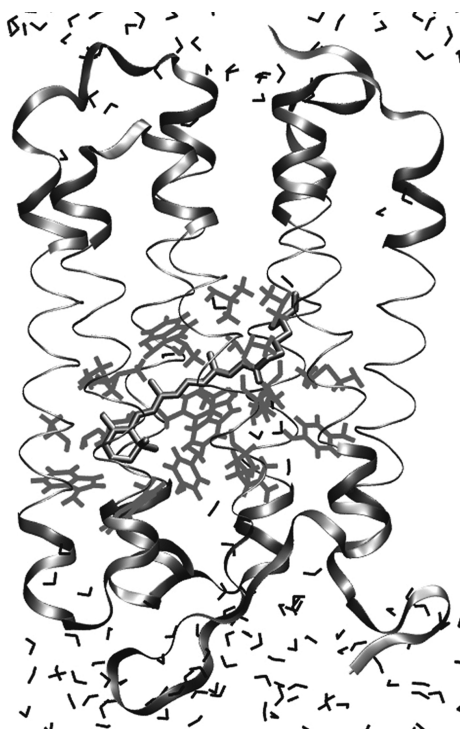
In a series of annealing MD simulations, in which all crystallographically determined coordinates of heavy atoms were kept fixed, the hydrogen and the other modeled atoms were allowed to adjust to the X-ray structure. As a result, for instance, the hydrogen bonded network shown in Figure 2 became apparent and the water molecules assumed the depicted orientations.

Because the charge distribution of BR polarizes the high-dielectric aqueous phases on the two sides of the purple membrane, the induced orientational polarization generates a reaction field acting back on the BR charge distribution. To provide a simple model for this reaction field, we attached two water caps comprising a total of 246 additional water molecules to the two water-exposed surfaces of BR. The caps were allowed to diffuse into the surfaces in a multiple stage computational relaxation process:

In a sequence of 300 K MD simulations, most protein atoms were allowed to move freely. Solely the positions  $x$  of the  $C_\alpha$  atoms and of the heavy atoms of the special residue LyR216 were constrained by harmonic potentials  $k_c(x - x_c)^2$  with the stiff force constants  $k_c = 6 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$  to the crystallographic positions  $x_c$ . These stiff constraints restrict thermal fluctuations to a root-mean-square deviation (rmsd) of about 0.3 Å and, thus, exclude large deviations from the X-ray structure at the backbone and at the chromophore. Furthermore, the added water molecules were constrained by soft harmonic potentials [ $k_c = 0.5 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ ], allowing fluctuations with an rmsd of about 1.1 Å to certain initial positions. These positions were given by the average positions in the respective preceding simulation (in contrast, the crystallographic water molecules were always bound by soft potentials to their X-ray positions). For the added water molecules, the resulting stationary positions served in all following MM–MD simulations as centers of constraining soft potentials. Here, most crystallographic water molecules remained softly constrained to the experimental positions. An exception were the members  $W_1$ – $W_3$  of the central water cluster, which were left unconstrained. The thus obtained structural model of BR preserves the key features of the original X-ray structure and, therefore, is called 1C3W/MD.

For the sake of computational efficiency, the MD simulation model 1C3W/MD contains as few atoms as possible and nevertheless, for the sake of accuracy, covers key aspects of the complicated and inhomogeneous dielectric environment represented by a membrane/water system. It is well-known that a model of a membrane protein, which omits the membrane and most of the surrounding water for reasons of computational efficiency, can be stable in extended MD simulations only if at least part of the system is constrained at the experimental





**Figure 3.** BR model for partially constrained MD simulations and for the computation of a polarized force field. The backbone of residues, whose  $C_{\alpha}$  atoms are constrained by stiff harmonic potentials to the crystallographic positions,<sup>40</sup> is drawn as a wide loop; the backbone of unconstrained residues is indicated by a thin circular drawing style. Water molecules are represented as small black hooks. The residues, whose partial charges are calculated by DFT/MM to generate a polarized force field, are drawn as light-gray structural models.

positions.<sup>65</sup> Because we are solely interested in the thermal fluctuations of all those BR atoms, which are located near the chromophore within the core of the membrane, we constrained the  $C_{\alpha}$  atoms of residues near the extra- and intracellular surfaces by stiff harmonic potentials to the crystallographic positions.

As shown in Figure 3, the unconstrained part of the simulation system forms a slab in the center of and oriented parallel to the membrane. The slab covers at least four windings of each of the seven transmembrane helices. Thus, its thickness along the membrane normal is everywhere larger than 22 Å. In its center, the slab contains the chromophore, whose extension along the membrane normal is about 8 Å.

The specification of the simulation setup is complete as soon as also a MM force field for the BR chromophore is given, because protein force fields like CHARMM22 solely cover parameters for standard molecular components of proteins. Thus, the parameters of the Lennard-Jones potentials for the various atom types in RSBH<sup>+</sup> were adopted from CHARMM22 parameters of atom types occurring in similar structural motifs. By the same analogy modeling, also the force constants of the harmonic bond stretch and angle deformation potentials were chosen.<sup>39</sup> In contrast, the equilibrium values of the bond lengths and angles entering these potentials were taken as averages from a series of DFT/MM calculations of the RSBH<sup>+</sup> embedded in the BR binding pocket. Also, the partial charge distribution within the RSBH<sup>+</sup> was extracted from these calculations. The details of these calculations are given further below. The parameters of the torsional potentials for the dihedral angles within the polyene chain were estimated from the quantum chemical results on RSBH<sup>+</sup> model chromophores and the accompanying analysis of experimental data in ref 66. These parameters are listed in Table 1. For the remaining torsional

**TABLE 1: Heights  $\Delta E$  of Torsional Potentials in BR's RSBH<sup>+</sup><sup>a</sup>**

double bonds		single bonds	
bond	$\Delta E^{\circ}$	bond	$\Delta E^{\circ}$
C5=C6	32	C6—C7	5
C7=C8	28	C8—C9	8
C9=C10	26	C10—C11	12
C11=C12	23	C12—C13	16
C13=C14	20	C14—C15	20
C15=N	30		

<sup>a</sup> % energies are given in kcal/mol.

potentials, which are required to define MM models for the cyclohexene ring, for the methyl groups attached to the polyene chain, and for the lysine chain, parameters were chosen which in CHARMM22 are standard for such structural motifs.<sup>39</sup>

Among all of these force field parameters, there are only a few which are of key importance for MM—MD simulations and all of them pertain to the conjugated chain within BR's RSBH<sup>+</sup>. These parameters are the partial charges, which sensitively depend on the electric field within the binding pocket and steer the important electrostatic interactions with the surroundings, the torsional potentials, which steer the possible deviations of the RSBH<sup>+</sup> from planarity, and the equilibrium values of the bond angles, which steer the banana-shaped curvature of the RSBH<sup>+</sup> (for the importance of the torsional potentials, see refs 66–68). As explained above, these key parameters were either derived from *in situ* DFT/MM descriptions or from a previous quantum chemical analysis.<sup>66</sup>

**MM—MD Simulation Procedures.** The MM—MD simulations were carried out with the program package EGO/MMII,<sup>69</sup> which combines a hierarchy of fast structure-adapted multipole expansions for the efficient treatment of the long-range electrostatics<sup>70,71</sup> with a multiple time-step integrator.<sup>72,73</sup> A period of 1 fs was chosen as the basic time step of this integrator. The bond lengths of the hydrogen atoms were kept constant using the M-SHAKE algorithm<sup>74</sup> with a relative tolerance of  $10^{-6}$ . The temperature of the system was kept at 300 K by coupling one Berendsen thermostat<sup>75</sup> to the protein atoms and another one to the water molecules (with the exception of the three water molecules near the RSBH<sup>+</sup>). As time constants of thermostat couplings, we chose 0.5 ps. Two separate thermostats were applied to the protein and the water subsystems to avoid the so-called “hot-solvent/cold-solute” problem.<sup>76</sup>

Using different force fields for the apoprotein, we carried out several MD simulations at 300 K with durations in the range 10–50 ns. In all of these simulations, the general starting model 1C3W/MD, in which all  $C_{\alpha}$  atoms are still stiffly constrained to the crystallographic positions, was converted into the partially unconstrained simulation model sketched in Figure 3 by a 150 ps equilibration simulation during which the constraints at the  $C_{\alpha}$  atoms in the central slab were removed and the system was tempered to 300 K. A second 1 ns equilibration phase preceded the trajectories used for data acquisition.

The force fields employed in these MD simulations differed in the partial charges assigned to the 23 residues forming the binding pocket of the RSBH<sup>+</sup>. We either chose the standard CHARMM22 partial charges<sup>45</sup> or partial charges obtained by iteratively applying the DFT/MM hybrid method of Eichinger et al.<sup>28</sup> to certain BR structures. The resulting sets of partial charges then define various structure-adapted polarized force fields. Following ref 77, one could call the approach applied by us to the computation of an improved electrostatic potential within BR also a partial “moving domain” DFT/MM method,

which is partial because it has been solely applied to the central region of BR containing the chromophore and its neighborhood (for a general discussion, we refer to ref 77).

**Calculation of Polarized Force Fields.** One of the BR structures chosen for the computation of a polarized force field (called PBR) was our general starting model 1C3W/MD. For the DFT/MM hybrid computation of partial charges, all 23 residues were chosen, which in 1C3W/MD have at least one atom at a distance smaller than 4 Å from the RSBH<sup>+</sup>. Each DFT/MM computation requires a partition of the whole structure 1C3W/MD into a DFT and a MM fragment. The partition is defined by specifying those C–C single bonds, at which the covalent linkage between the two fragments is cut using the “scaled position link atom method” (SPLAM).<sup>28</sup>

By cutting for most of the selected residues (Met20, Val49, Ala53, Tyr57, Tyr83, Trp86, Thr89, Thr90, Leu93, Met118, Trp138, Ser141, Thr142, Met145, Trp182, Tyr185, Pro186, Trp189, and Ala215) the C<sub>α</sub>–C<sub>β</sub> bond, the DFT treatment was restricted to the respective side chain. For the four charged residues Asp85, Asp212, Arg82, and Lyr216 in the center of BR, however, also a piece of the backbone was included. Here, for the respective residue *i*, two SPLAM cuts were applied to the C<sub>α,i-1</sub>–C<sub>i-1</sub> and C<sub>α,i</sub>–C<sub>i</sub> bonds.

The DFT calculations were carried out with the plane-wave code CPMD<sup>78</sup> using the gradient-corrected exchange functional of Becke,<sup>79</sup> the correlation functional of Perdew,<sup>80</sup> and the norm-conserving pseudopotentials of Troullier and Martins.<sup>81</sup> The rectangular box containing the grid for the plane-wave expansion of the Kohn–Sham orbitals was placed around the respective DFT fragment in such a way that no atom of the fragment came closer than 3 Å to one of the faces and that the box volume became minimal. The size of the basis set was then defined by a cutoff of 70 Ry. The electrostatic potential generated by the MM fragment was imported into the respective DFT Hamiltonian as described in ref 28. So-called “ESP” partial charges were calculated for the DFT fragment by optimizing the match with the electrostatic potential, which is due to the enclosed charge distribution, at the van der Waals surface of the fragment.<sup>82</sup> After each DFT/MM calculation on one of the residues, its previous partial charges were replaced by these new ESP charges. This procedure, which started at the CHARMM22 partial charges, was iterated several times over the whole list of selected residues until the ESP charge distribution became stationary. Note that in each cycle the DFT/MM calculations started at Lyr216 up to the last cycle which also ended with Lyr216.

The thus obtained ESP charges, which characterize the polarized force field PBR, are listed for the 23 residues in Tables 2–9 provided online as Supporting Information (SI). In addition to our X-ray-based starting model 1C3W/MD, we applied this DFT/MM approach for computing a polarized force field to several further structural models, which were obtained as snapshots from MD simulations (see the Results and Discussion).

**Generation of Free Energy Maps.** As indicators, to what extent the chromophore binding pocket maintains its structure during the MD simulations, we used the two conformational coordinates *d<sub>i</sub>*, *i* = 1, 2, defined in Figure 2 and in the associated caption. For the computation of the *d<sub>i</sub>*, we extracted from each MD trajectory every 500 fs the distances δ<sub>α</sub>(*t*) between all of those oxygen atoms α and the Schiff base proton, which are possible hydrogen bonding partners of the RSBH<sup>+</sup>. The atoms α comprise the oxygens of the three water molecules (α = 1, 2, 3) in the binding pocket and the two oxygens (α = 4, 5) of Asp85. The trajectories δ<sub>α</sub>(*t*) obtained in the various simulations

are documented either in the Results and Discussion or in the SI. By the nearest neighbor criterion explained in the caption to Figure 2, the trajectories δ<sub>α</sub>(*t*) with α = 1, 2, 3 were reduced to a data set *d<sub>1</sub>*(*t*) and those with α = 4, 5 to a set *d<sub>2</sub>*(*t*).

The data sets *d<sub>i</sub>*(*t*) were converted into free energy landscapes Δ*G*(*d<sub>1</sub>*, *d<sub>2</sub>*). For this purpose, the rectangle [1.25 Å, 4.5 Å] × [1.25 Å, 4.25 Å] within the (*d<sub>1</sub>*, *d<sub>2</sub>*)-plane was subdivided into 31 × 33 bins and histogram density estimates *p*(*d<sub>1</sub>*, *d<sub>2</sub>*) were evaluated from the distance data sets. The free energy then is Δ*G*(*d<sub>1</sub>*, *d<sub>2</sub>*) = –*k<sub>B</sub>T* ln *p*(*d<sub>1</sub>*, *d<sub>2</sub>*) – *G*<sub>0</sub>, where *k<sub>B</sub>* is the Boltzmann constant and *G*<sub>0</sub> is a constant serving to set the minimum of Δ*G* to zero. Because empty bins in the histogram density estimate [*p*(*d<sub>1</sub>*, *d<sub>2</sub>*) = 0] would lead to infinite free energies, an upper energy cutoff Δ*G*<sub>max</sub> = –*k<sub>B</sub>T* ln(1/*M*<sub>max</sub>) was introduced, where *M*<sub>max</sub> is the maximum bin count found in the histogram, and Δ*G* was set to Δ*G*<sub>max</sub> at all empty bins. A nearest neighbor smoothing was applied before generating the contour plots of Δ*G*(*d<sub>1</sub>*, *d<sub>2</sub>*).

**Conformational Analysis of the Lysine Chain.** To explore the equilibrium conformational ensemble of the aliphatic chain connecting in Lyr216 the RSBH<sup>+</sup> with the backbone, we have used a Hamiltonian replica exchange (HRE) approach.<sup>83–85</sup> Like in conventional temperature replica exchange,<sup>86,87</sup> the Metropolis criterion

$$P_{ij} = \min[1, \exp(-\Delta_{ij})] \quad (1)$$

controls the exchange of two replicas *i* and *j* at neighboring temperatures *T<sub>i</sub>* < *T<sub>j</sub>*. In HRE, however, the difference Δ<sub>*ij*</sub> has to account additionally for different Hamiltonians *H<sub>i</sub>*(*X*) and *H<sub>j</sub>*(*X*) of the replicas at the temperatures *T<sub>i</sub>* and *T<sub>j</sub>*:

$$\Delta_{ij} = \beta_i[H_i(X_j) - H_i(X_i)] + \beta_j[H_j(X_i) - H_j(X_j)] \quad (2)$$

Here, *X<sub>i</sub>* and *X<sub>j</sub>* denote the configurations of the replicas before exchange, β is (1/*k<sub>B</sub>T*), and *k<sub>B</sub>* is the Boltzmann constant.

Our HRE approach closely resembles the replica exchange with solute tempering (REST) method introduced by Liu et al.<sup>85</sup> The differences pertain to the partitioning of the simulation system. Whereas in REST a peptide is heated and the surrounding solvent remains “cool” through scaling of the Hamiltonian, in our HRE, the lysine chain from C<sub>β</sub> to C<sub>ε</sub> is heated and the remainder of the simulation system stays “cool” upon scaling.

For a more specific explanation, we denote the lysine–lysine, lysine–remainder, and remainder–remainder interactions by L, LR, and R distinguishing additionally some of the bonded (b) and nonbonded (nb) interactions. We choose for a temperature *T* and the sampling temperature *T*<sub>0</sub> the Hamiltonian *H* by

$$H = H_L + H_{LR}^b + \sqrt{\frac{T}{T_0}} H_{LR}^{nb} + \frac{T}{T_0} H_R \quad (3)$$

This choice reduces eq 2 to

$$\Delta_{ij} = (\beta_j - \beta_i) \{ [H_L(X_i) + H_{LR}^b(X_i)] - [H_L(X_j) + H_{LR}^b(X_j)] \} + (\sqrt{\beta_j T_0} - \sqrt{\beta_i T_0}) [H_{LR}^{nb}(X_i) - H_{LR}^{nb}(X_j)] \quad (4)$$

This expression shows that the difference Δ<sub>*ij*</sub> entering the exchange probability (1) is calculated from the potential energy of the lysine chain (including its interactions with the remainder), because the much larger term *H<sub>R</sub>* cancels. The small number *N*

of degrees of freedom in the lysine chain leads to smaller shifts of the potential energies at neighboring temperatures such that the exchange criterion is frequently met even for large temperature distances. As a result, HRE allows us to cover the temperature range from 300 up to 660 K with only five replicas ( $T_i \in \{300, 360, 440, 540, 660 \text{ K}\}$ ) and an average exchange probability  $P_{ij}$  between neighboring replicas ( $j = i \pm 1$ ) of about 40%. The HRE sampling was extended to 30 ns, and exchanges were attempted every 5 ps.

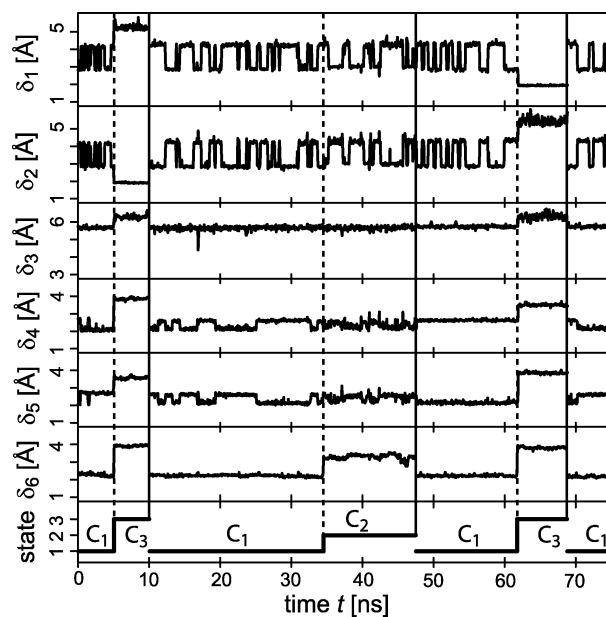
To identify the conformations of the lysine chain, we applied the hierarchical clustering approach described by Carstens et al.<sup>88</sup> in a paragraph entitled "Determination of the Conformational Ensembles". In short, we extracted from the HRE trajectory at the target temperature  $T_0 = 300 \text{ K}$  every 5 ps the five torsional angles  $\psi_k(t)$ ,  $k = 1, \dots, 5$ , which belong to the five single bonds connecting the Lys216  $C_\alpha$  atom with the N atom of the RSBH<sup>+</sup> (cf. Figure 1). The resulting data set  $\mathcal{X}$  covered 6000 five-dimensional vectors  $\Psi(t)$  in the cyclic space  $[0, 2\pi]^5$ . Using a mixture of 100 five-dimensional, univariate, and cyclic normal distributions with equal weights and widths, a maximum likelihood estimate  $\rho(\Psi)$  was calculated for the data density in  $\mathcal{X}$  by applying the algorithms suggested in refs 89 and 90. Convolution of  $\rho(\Psi)$  with Gaussians of varying widths yielded a scale space of successively smoothed model densities  $\rho(\Psi|\sigma)$ , whose maxima indicate for each resolution scale  $\sigma$  the prototypical conformations. The SI documents the details of this hierarchical conformational analysis, which safely identifies the minima of free energy landscapes in high-dimensional spaces at varying resolutions.

## Results and Discussion

Here, we first demonstrate that our partially constrained MD simulation model for BR (cf. Figure 3) exhibits, for a conventional nonpolarizable force field, the same type of structural instability and decay into the non-native conformation  $C_{nn}$ , which was previously observed with such force fields by Kandt et al.<sup>50</sup> for an unconstrained protein trimer in a huge membrane/water system and by Eichinger<sup>39</sup> and Hayashi et al.<sup>55</sup> for partially constrained monomeric models.

**CHARMM Instability of 1C3W/MD.** To analyze the equilibrium fluctuations, which are predicted by the CHARMM22 force field for our BR model, we have carried out four MD simulations at 300 K covering a total of 75 ns as described in Methods. Figure 4 depicts the resulting trajectories for six selected distances  $\delta_\alpha$ ,  $\alpha = 1, \dots, 6$ , capable of indicating the stability of the chromophore binding pocket. In the bottom panel, the figure additionally displays the trajectory of a classification, which assigns to each sampled configuration of the aliphatic chain connecting the RSBH<sup>+</sup> with the protein backbone one of three conformational states  $C_k$  (this classification will be explained further below). The trajectories were sampled every 500 fs. The resulting data sets were smoothed and reduced by combining 100 consecutive data points into 50 ps averages.

Each of the four CHARMM22 simulations started at the initial structure 1C3W/MD (in Figure 4, these simulations are separated by vertical solid lines). In each simulation, the protein was found (bottom panel) to relax immediately into a first conformational state. Here, LyR216 maintains its initial structure, which is characterized by the conformation  $C_1$  of the aliphatic chain and by the backbone distance  $\delta_6 = 2.1 \text{ \AA}$ . The latter value shows that the backbone segment of LyR216 is stably H-bonded to the amide group of Gly220 in the crystallographic  $\alpha$ -helical arrangement. Note in this context that LyR216 is located at the beginning of the C-terminal part of  $\alpha$ -helix G just after the



**Figure 4.** Trajectories of six distances  $\delta_\alpha(t)$ ,  $\alpha = 1, \dots, 6$ , and of a classification of the lysine chain into three conformational states  $C_1$ ,  $C_2$ , and  $C_3$  obtained from four CHARMM22-MD simulations at 300 K each starting at the structure 1C3W/MD (total duration: 75 ns). The first five  $\delta_\alpha$  measure the distances between the Schiff base proton and the five oxygen atoms, which are possible hydrogen bonding partners ( $\delta_1$ – $\delta_3$  for the water molecules  $W_1$ – $W_3$  and  $\delta_4$  and  $\delta_5$  for Asp85).  $\delta_6$  is the distance between the backbone carbonyl oxygen of LyR216 and the amide hydrogen of Gly220. For a discussion, see the text.

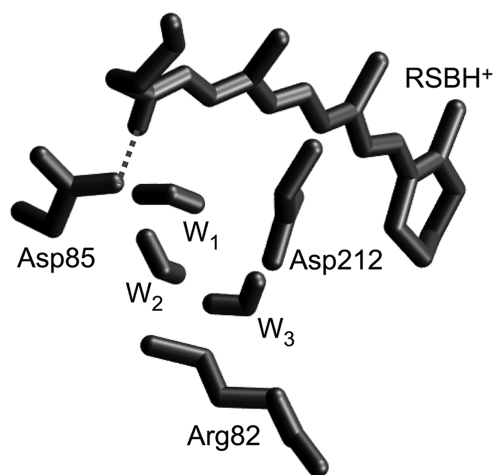
$\pi$ -bulge at Ala215, which separates the two differently oriented parts of  $\alpha$ -helix G.<sup>40</sup>

In the first three simulations, a conformational transition (marked by the dashed lines in Figure 4) occurred after 5, 25, and 15 ns, respectively. In these transitions, the initial conformation  $C_1$  of the aliphatic chain in LyR216 changed into conformation  $C_2$  or  $C_3$ . Correspondingly, in Figure 4, the distance  $\delta_6(t)$  increases to 3.3 Å ( $C_2$ ) or 3.8 Å ( $C_3$ ), indicating that the associated H-bond is broken, i.e., that LyR216 has abandoned its original integration into  $\alpha$ -helix G. The reverse transition back to  $C_1$  was never observed. Thus, on a time scale of about 10 ns, the CHARMM22 model of BR predicts conformational transitions involving a partial unzipping of  $\alpha$ -helix G.

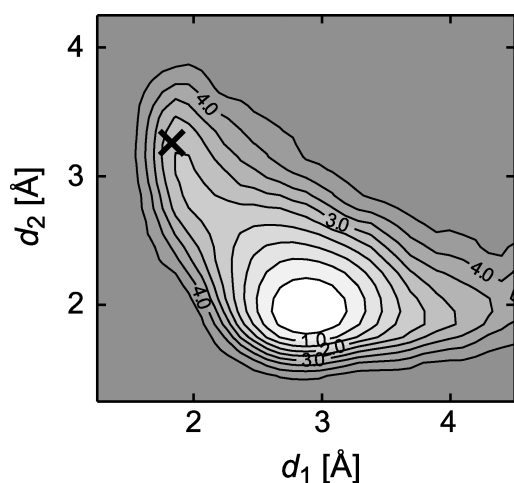
A closer look at the first five distance trajectories  $\delta_\alpha(t)$  in Figure 4 reveals that the CHARMM22 model of BR deviates from 1C3W in further structural details. These deviations show up on a much shorter time scale, i.e., during the initial relaxation into the conformation  $C_1$ . In  $C_1$ , the trajectories  $\delta_4(t)$  and  $\delta_5(t)$  are strictly anticorrelated, switching between values of 2.1 and 2.7 Å, respectively. These values show that the RSBH<sup>+</sup> is always H-bonded to one of the Asp85 oxygens. The trajectories indicate that the two oxygens exchange place about every 4 ns. Correspondingly, in  $C_1$ , the water molecules are found at much larger distances with  $W_3$  remaining stably at 5.7 Å.  $W_1$  and  $W_2$  are seen to exchange every 1.2 ns their positions at distances of 2.8 and 4.3 Å from the Schiff base proton. In conformation  $C_2$ , the considered trajectories  $\delta_\alpha(t)$  are similar except for a slightly larger distance (3.0 Å instead of 2.8 Å) of the closest water molecule. For  $C_3$ , however, the data in Figure 4 indicate a stable H-bond to either  $W_1$  or  $W_2$ .

In summary, the CHARMM22 model of BR predicts either a decay of the stable H-bonded network displayed by Figure 2 (for  $C_1$  and  $C_2$ ) or a partial decay of the  $\alpha$ -helix G (for  $C_3$ ) or





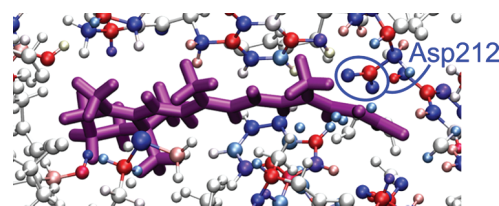
**Figure 5.** Typical structure of the chromophore binding pocket during those 50 ns of MD simulation with the CHARMM22 force field, during which the crystallographic conformation  $C_1$  of LyR216 was maintained. The structure features an H-bond (gray dashed line) between Asp85 and the protonated Schiff base and is typical for the non-native conformation  $C_{nn}$  of the hydrogen bonded network.



**Figure 6.** CHARMM22 free energy landscape  $\Delta G(d_1, d_2)$  (in kcal/mol) for the conformation  $C_1$  of the lysine chain obtained from 50 ns of MD simulation at 300 K.  $\Delta G$  is a function of the distances between the Schiff base proton and the nearest oxygen of a water molecule ( $d_1$ ) or of Asp85 ( $d_2$ ). The black cross marks the values of these distances in the all-atom model 1C3W/MD of the crystallographic structure.<sup>40</sup>

both (for  $C_2$ ). Figure 5 shows a typical snapshot from a CHARMM22 trajectory with LyR216 in the crystallographic conformation  $C_1$ . As expected above for a simulation with a nonpolarizable force field, the stable hydrogen bonded network displayed by Figure 2 has been replaced by a stable H-bond between the RSBH<sup>+</sup> and Asp85, characterizing the conformational state  $C_{nn}$ . Due to the strong association of the RSBH<sup>+</sup> with a single counterion and the well-known sensitivity of the RSBH<sup>+</sup> to counterion distance, such structures are invalid as starting points for DFT/MM calculations of the chromophore's IR spectra.<sup>39</sup>

The free energy landscape  $\Delta G(d_1, d_2)$  depicted in Figure 6 provides a measure for the strength of the association between the RSBH<sup>+</sup> and Asp85 in the CHARMM22 simulations with lysine conformation  $C_1$ .  $\Delta G(d_1, d_2)$  has been obtained from the first five trajectories  $\delta_\alpha(t)$ , whose smoothed versions are depicted in Figure 4, as described in Methods. According to Figure 6,  $\Delta G(d_1, d_2)$  has a minimum at the H-bonding distance  $d_2 \approx 2$  Å between Asp85 and the Schiff base proton which is characteristic for the non-native conformation  $C_{nn}$  of the hydrogen bonded



**Figure 7.** Polarization charges  $Q_{\text{pol}} - Q_{\text{CHARMM}}$  in the chromophore binding pocket. Negative polarization charges are drawn in red, positive ones in blue (magenta: RSBH<sup>+</sup>).

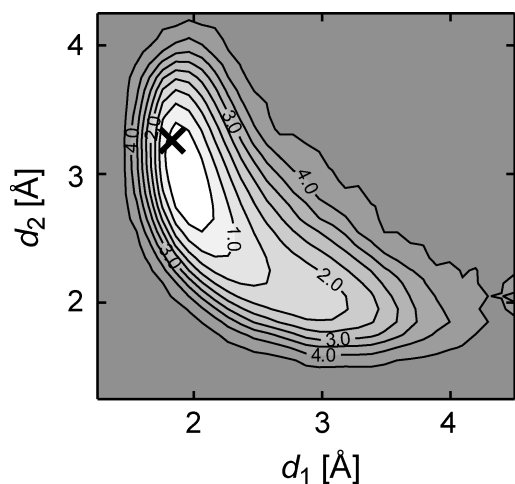
network. In contrast, the all-atom model 1C3W/MD (black cross in Figure 6) assigns a much larger value (3.3 Å) to  $d_2$  and signifies the H-bond between the Schiff base and a water molecule by the small value  $d_1 \approx 1.9$  Å. At the position of the cross, the free energy  $\Delta G$  is about 3 kcal/mol above the minimum, indicating that the crystallographic structure represents only a rare fluctuation in MD simulations of 1C3W/MD with the CHARMM22 force field (note here once again that the crystallographic structure 1C3W/MD features the lysine conformation  $C_1$  and a stable hydrogen bond between LyR216 and Gly220).

**PBR Stability of 1C3W/MD.** To check whether the use of a polarized force field, which should provide an improved description of the electrostatic potential within BR, can remove the CHARMM22 instability described above, we applied the iterative DFT/MM procedure outlined in Methods to the 23 residues making up the binding pocket of the BR chromophore. The root-mean-square deviation (rmsd) between the partial charges of the resulting PBR force field and the standard CHARMM22 charges was about a tenth of an elementary charge (0.13  $e$ ). Much larger ( $>0.2$   $e$ ) than this 23 residue average were the local rmsd's for Asp85, Asp212, and Leu93. Still, sizable rmsd's were found at Met20, Met118, Tyr83, and Tyr57 (for detailed comparisons, see Tables 2–9 in the SI).

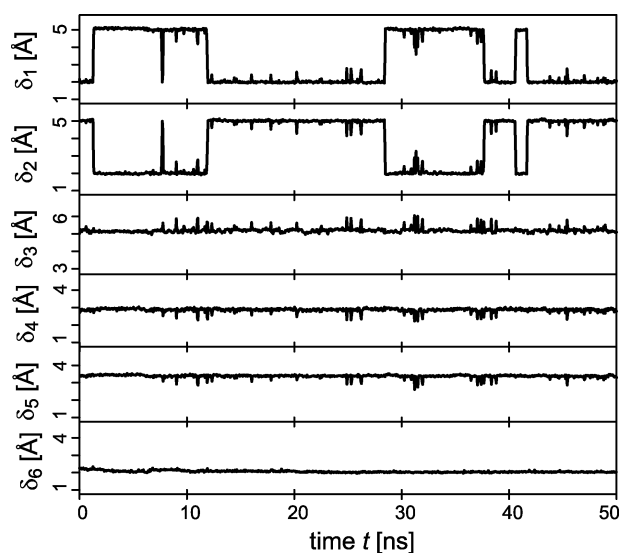
Figure 7 illustrates these results by highlighting the differences between the PBR and CHARMM22 partial charges. Considering specifically Asp212, one recognizes that additional negative charge is attracted by the positively charged chromophore and is shifted from the  $C_\gamma$  atom to the two terminal oxygen atoms. Thus, at Asp212, the PBR charge distribution differs from the CHARMM22 standard by an additional induced dipole pointing (as expected) away from the RSBH<sup>+</sup> for its improved solvation. A similar geometry of induced dipoles is also observed at other residues.

Figure 8 shows the free energy landscape  $\Delta G(d_1, d_2)$  obtained from a 50 ns simulation with the polarized force field PBR. Comparing this landscape visually with the CHARMM22 result in Figure 6, one immediately recognizes that the experimental H-bonded structure of the chromophore binding pocket (Figure 2) has been preserved in the PBR simulation. Now the minimum of  $\Delta G(d_1, d_2)$  is found quite close to the experimental (low-temperature) structure indicated by the cross. Thus, the H-bond between the protonated Schiff base and a water molecule has apparently been quite stable during this extended MD simulation.

A look at Figure 9, which displays the trajectories  $\delta_\alpha(t)$  of the six distances employed as measures for conformational stability (cf. Figure 4), immediately proves that the H-bonded network in the chromophore binding pocket remained rock-solid during the 50 ns of MD simulation with PBR. The only apparent dynamics is an occasional (every 6 ns) exchange of the water molecules  $W_1$  and  $W_2$  as hydrogen bonding partners of the RSBH<sup>+</sup>. The distance  $\delta_6$  remains constantly small at 2.1 Å, proving that at LyR216 the  $\alpha$ -helical structure of helix G was preserved and indicating that the lysine chain did not undergo



**Figure 8.** Free energy landscape  $\Delta G(d_1, d_2)$  for the polarized force field PBR. For further explanation, see the text and the caption to Figure 6.



**Figure 9.** Trajectories of six distances  $\delta_\alpha(t)$ , from a 50 ns PBR-MD simulation at 300 K starting at the structure 1C3W/MD. For further explanations, see the caption to Figure 4.

any conformational transitions. In fact, a conformational classification of the torsional angles within the lysine chain has demonstrated that the initial conformation  $C_1$  was maintained during the 50 ns of simulation (data not shown).

The above data have demonstrated that the electronic polarization of the RSBH<sup>+</sup> and of the surrounding residues can stabilize the chromophore binding pocket. For reasons discussed further above, we employed in the PBR simulation the standard nonpolarizable TIP3P model potential for the water molecules, thus neglecting possible polarization effects at these sites. To obtain an estimate on the size of these effects, we have applied our DFT/MM approach (after completion of the above MD study) also to the three water molecules in the binding pocket for the structure 1C3W/MD. Independently of the force field employed for the apoprotein (PBR or CHARMM22) and hardly varying with position and orientation (cf. Figure 2), we found that the dipole moments of these molecules were about 2.59 D instead of the 2.35 D characteristic for TIP3P. This value of 2.59 D is also larger than the average dipole moment of 2.41 D determined previously by the same DFT/MM approach for DFT water molecules in liquid MM water environments.<sup>35</sup>

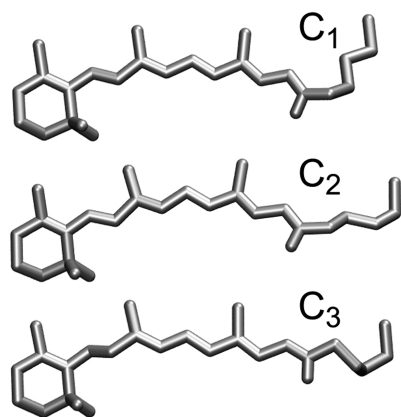
This result came unexpected to us. In view of the strong local electric fields acting in bulk water on a given molecule, we had

assumed that the mean field polarization coded into a bulk water model like TIP3P should suffice to describe also the polarization of a few molecules in a hydrophobic environment. However, within a nonpolar environment near the center of a membrane and within the rigid and tightly bound network characterizing the binding pocket, the local electric fields are apparently stronger than expected. Thus, for the water molecules, our PBR simulation has underestimated the energetic stability of the associated H-bond interactions by about 10%, implying that the H-bonded network of Figure 2 is even more stable than suggested by Figure 8. An improved estimate on the size of this additional stabilization would have required lengthy MD simulations with a correspondingly extended polarized force field covering water molecules with larger dipole moments. However, because the stabilization of the experimental BR structure through a polarized force field has already been demonstrated and because such simulations are associated with a considerable computational cost, we did not address this issue further.

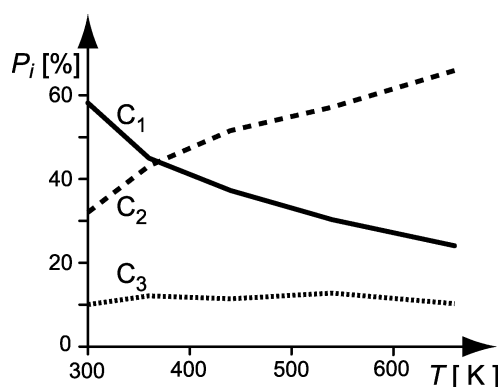
**Heterogeneity.** The 50 ns PBR simulation did not show any signs of conformational instability (neither for the hydrogen bonded network nor for the lysine chain nor for helix G). Thus, the conformational transitions and the new conformational states observed in three of the four CHARMM22 simulations could have simply been an artifact of the CHARMM22 force field. On the other hand, one of the many DFT/MM geometry optimizations of the RSBH<sup>+</sup>, which started at a snapshot from the PBR-MD trajectory and was carried out for the purpose of computing a vibrational spectrum [see part II of this work (DOI 10.1021/jp902432e)], had accidentally also shown a transition of the lysine chain into a different conformation. Therefore, we were not sure as to whether other conformational states of the lysine chain are accessible at room temperature in BR.

However, if other conformational substates should actually exist in the 300 K equilibrium ensemble, then we had to expect from our 50 ns PBR simulation that they are separated by substantial free energy barriers from the crystallographic conformation. Because of the height of these barriers, a straightforward extension of simulation times offered no chance for a valid estimate of their possible contribution and alternative sampling techniques were required. Conventional replica exchange techniques<sup>86,87</sup> are computationally very costly and cannot guarantee the convergence of the sampling within the necessarily limited simulation times (see, e.g., ref 91 for a discussion). Therefore, we designed a Hamiltonian replica exchange (HRE) technique that concentrates on the few degrees of freedom within the lysine chain and therefore opens the chance that the 300 K equilibrium ensemble of lysine configurations can be sampled (see Methods).

The 30 ns of HRE sampling yielded a 300 K ensemble of lysine configurations, whose structure was analyzed by the hierarchical clustering procedures described in Methods. The SI presents details of the corresponding results for various levels of resolution. Here, we restrict our attention to a coarse level of resolution, at which the generalized HRE ensemble of lysine configurations is characterized by three main conformational states  $C_i$ ,  $i \in \{1, 2, 3\}$ . These states are characterized by certain combinations  $\Psi_i \in [0, 2\pi]^5$  of the five torsional angles within the lysine chain. For the recovery of the corresponding complete structures, the 300 K trajectories  $\Psi(t)$  were scanned for most similar points  $\Psi(t_i)$  and the snapshots at these time points  $t_i$  were chosen as prototypical structures. Figure 10 shows for each of the three main conformations the thus obtained prototypical



**Figure 10.** The three conformations of LyR216 dominating the PBR equilibrium ensemble at 300 K as determined by a hierarchical cluster analysis.  $C_1$  is the conformation found in the X-ray structure 1C3W, and conformations  $C_2$  and  $C_3$  are alternate conformations previously observed in the 300 K simulation with the CHARMM22 force field (cf. Figure 4). The drawings cover the heavy atoms (and the Schiff base proton) within the LyR216 side chain.



**Figure 11.** Occupancies  $P_i$  of the three conformational states  $C_i$  determined in generalized ensemble at the five replica temperatures  $T \in \{300, 360, 440, 540, 660\}$  K during the last 25 ns of HRE sampling.

structures of residue LyR216 (note that we have used this three-state classifier repeatedly in the previous sections).

In the generalized HRE ensemble, the conformational states  $C_i$  occur at different relative frequencies  $P_i$ , which depend on the temperature  $T$  of the respective replica. Although in the HRE setting only the replica at the target temperature  $T_0 = 300$  K is of key physical importance, it is nevertheless instructive to consider the temperature dependence of the state occupancies  $P_i(T)$ . Figure 11 shows the corresponding data, which were derived from the last 25 ns of HRE sampling. The first 5 ns were considered as a relaxation phase because all replicas had been initialized at the model 1C3W/MD belonging to conformation  $C_1$ . At  $T_0$ , state  $C_1$  is dominant (58%),  $C_2$  represents a considerable population  $P_i$  (32%), and  $C_3$  a minor (10%) population  $P_i$ . Interestingly, the populations  $P_i$  of the two frequent conformations  $C_1$  and  $C_2$  show an opposite temperature dependence:  $P_1$  decreases whereas  $P_2$  increases with  $T$ . Thus, it seems that  $C_1$  is enthalpically and  $C_2$  entropically favored.

This interpretation agrees with further observations. (i) The hierarchical classification tree depicted in Figure 15 of the SI shows a strong splitting of the class  $C_2$  into many substates at finer resolutions. In contrast, the class  $C_1$  does not show a similarly strong splitting. (ii) The average extension of the lysine chain differs in the various states. From the PBR ensemble at 300 K, we extracted for the distance  $\delta_7$  between the Schiff base

nitrogen and the  $C_\alpha$  atom of LyR216 the values 5.6 Å ( $C_1$ ), 5.4 Å ( $C_2$ ), and 4.7 Å ( $C_3$ ). Applying the concepts of entropic springs, the most extended (crystallographic) conformation  $C_1$  is likely to be compatible with fewer configurations of the aliphatic chain than the shorter conformation  $C_2$  and, therefore, should represent a lower entropy state. Note that at very short distances  $\delta_7$  the number of thermally accessible configurations is expected to decrease once again, as seems to be the case for  $C_3$ . (iii) The shorter length  $\delta_7$  of the lysine chain in conformations  $C_2$  and  $C_3$  is always associated with an elongated distance  $\delta_6$ , indicating a broken  $\alpha$ -helical H-bond between LyR216 and Gly220 (HRE data for  $\delta_6$  not shown, but see Figure 4 and the SI). Thus, the intact H-bond in  $C_1$  should enthalpically favor this state.

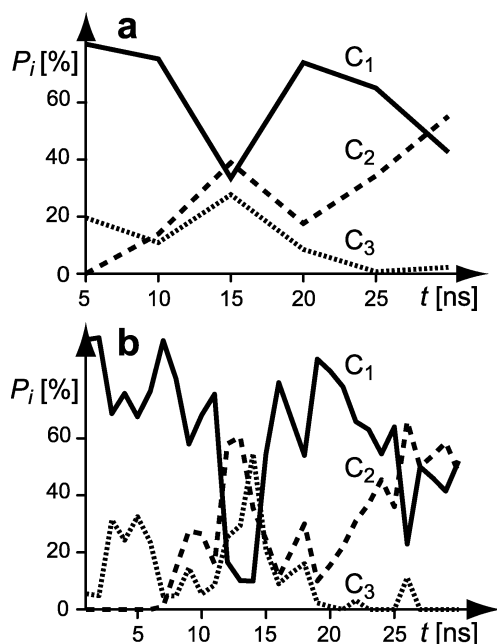
## Discussion

For BR, the HRE results predict that the (crystallographic) state  $C_1$  is, in fact, dominantly populated at the low temperatures (100 K) of the X-ray data collection and that other conformations will become increasingly important at physiological temperatures. To check this issue, we have applied our classifier to the 12 room-temperature NMR structures, which are considered as the best refinement results for the all-*trans* RSBH<sup>+</sup>.<sup>48</sup> We found that three of these NMR structures are classified as  $C_1$  and the remaining nine structures as  $C_2$ . In view of the fact that the NMR refinement employs an MD-based simulated annealing procedure, the dominance of the entropically favored high-temperature conformation  $C_2$  in the NMR ensemble is no surprise (for further details, see the SI). However, independent of this detail, the ensemble of refined NMR structures has thus shown that the NMR data are compatible with the lysine conformations  $C_1$  and  $C_2$ . Therefore, these data are also compatible with the results of our HRE simulations.

From the conformational analysis of the NMR results, one then expects that the NMR structures should show the broken  $\alpha$ -helical H-bond at LyR216, which the MD simulations consistently predict as a correlate of conformation  $C_2$ . Unfortunately the NMR structures cannot furnish such a result, because the H-bond distances within the helices had been restrained to the crystallographic values in the final phase of NMR refinement.<sup>48</sup> Thus, the NMR structures cannot answer the important question as to what extent the  $C_2$  conformation of the lysine chain is correlated with a melting of helix G near LyR216. Note in this connection that this local melting would be impeded, if helix G were  $\alpha$ -helically intact. However, because of the  $\pi$ -bulge at Ala215, helix G is kinked with LyR216 marking the start of a new  $\alpha$ -helical segment. It is well-known that melting disorder preferentially occurs at the ends of such segments. Physiologically, the  $\pi$ -bulge apparently creates an enhanced conformational flexibility for the covalent linkage between the chromophore and the RSBH<sup>+</sup>.

In summary, the HRE results have clearly shown that the two additional conformations of the lysine chain encountered in the CHARMM22 simulations are also elements of the 300 K ensemble belonging to the PBR force field. Furthermore, they have demonstrated that the populations of these conformational states are temperature dependent. Here, particularly the crystallographic conformation  $C_1$  is favored at low temperatures and the alternative conformation  $C_2$ , which dominates the NMR ensemble, is increasingly populated at higher (i.e., physiological) temperatures. Note at this point that the Schiff base stays stably hydrogen bonded to a water molecule in all alternative conformations as long as one employs the polarized force field PBR.





**Figure 12.** Temporal changes of the time-average occupancies  $\langle P_i \rangle_{\Delta t}(t_k)$  of the three conformational states  $C_i$  in the 300 K replica during the 30 ns of HRE. The occupancies are given at times  $t = k\Delta t$ ,  $k = 1, 2, \dots$ , with the averaging times (a)  $\Delta t = 5$  ns and (b)  $\Delta t = 1$  ns.

Two important questions remain. One pertains to the mean first passage times of transitions between the conformations and the other to the accuracy of our estimates on the populations  $P_i$  at physiological temperatures. Unfortunately, we cannot provide quantitative and statistically sound answers to these questions, because the associated computational cost is intractable for our means.

Concerning the transitions between the conformational states of the lysine chain, we solely can state that we have never observed any such transition within our 300 K PBR-MD simulations. Neither the 50 ns simulation of conformation  $C_1$  (cf. Figure 9) nor two further 20 ns simulations, which were initialized at conformations  $C_2$  and  $C_3$ , respectively, showed any sign of a conformational transition into a different state (the latter simulations are documented in the SI). Thus, the lifetimes of the lysine conformations  $C_i$  must be larger than about 0.1  $\mu$ s. However, these lifetimes may well be several orders of magnitude larger than that (e.g., several  $\mu$ s).

Concerning the question as to what extent the 30 ns HRE sampling has converged toward the equilibrium state of the generalized ensemble, one cannot give a definite answer. The reason for this uncertainty is rather subtle: If the conformational transitions were exclusively confined to the lysine chain (which was our working hypothesis), the huge number of transitions observed during the HRE sampling within that chain would guarantee convergence. However, as we have detected during our analysis of the HRE trajectories (see above), the transition from  $C_1$  to alternative conformations is always associated with a breaking of an H-bond in helix G. However, this helix was excluded from the rapid HRE sampling by our initial partitioning of the system into “hot” and “cool” regions and its conformational dynamics proceeds at the slow speed of a 300 K ensemble. Therefore, despite our extended sampling, the convergence cannot be guaranteed and valid error bounds for the populations  $P_i(T_0)$  cannot be given.

As an illustration for the difficulties of achieving convergence, consider Figure 12. At the coarse time resolution of 5 ns (Figure 12), it may seem as if state  $C_1$  tends to become less populated

with increasing HRE simulation time than state  $C_2$ . Furthermore, state  $C_3$  even seems to disappear. However, at the finer time resolution of 1 ns, the large sizes of the occupancy fluctuations become apparent, indicating that such conclusions are premature.

Concerning the sizes of the populations at 300 K, the HRE results thus solely prove that there are two similarly frequent conformations  $C_1$  and  $C_2$ , whereas other conformations occur at a much smaller frequency. Interestingly, the opposite behavior of the temperature dependences of  $C_1$  and  $C_2$  documented in Figure 11 is a much more stable result of the HRE simulations. Similar curves were obtained already when we had only a total of 20 ns simulation time for statistical analysis.

The conformational heterogeneity thus identified for the light-adapted state  $BR_{568}$  has important implications for the photocycle. Different initial conformations of the lysine chain will quite certainly lead to conformationally different photoproducts, because the photoisomerization involving the  $C_{13}=C_{14}$  bond is mechanically coupled to lysine chain motions. In line with these concepts and the above results, a conformational heterogeneity has been postulated already two decades ago by Diller and Stockburger<sup>92</sup> from room-temperature time-resolved resonance Raman spectroscopy. These authors interpreted their spectra of the  $L_{550}$  intermediate, which shows up a few  $\mu$ s after photoisomerization, in terms of two different conformational states for  $BR_{568}$  and  $L_{550}$  (called  $\alpha$  and  $\beta$ ). As relative occupancies, they estimated 58 and 42%, respectively. As key differences, they identified the kinetics of appearance and decay of the L intermediate for the two conformations ( $L_\alpha$  being fast and  $L_\beta$  slow).<sup>93</sup> Because of the statistical uncertainties in our calculation of relative populations (see above), we unfortunately cannot be sure, which of our conformations  $C_1$  and  $C_2$  belongs to  $\alpha$  and which to  $\beta$ .

In agreement with the concept of a conformational heterogeneity persisting up to the L state, a recent analysis of the BR photocycle kinetics has led Zimanyi et al.<sup>94</sup> to the conclusion that “no photocycle scheme with a single L intermediate is acceptable”. However, we do not share the strong conviction of Diller and Stockburger<sup>92</sup> that the conformational heterogeneity must persist also to the millisecond time scale of the photocycle (this view has been most recently reiterated by Hendler et al.<sup>95</sup> on the basis of yet another analysis of the photocycle kinetics). If the conformational transitions within the lysine chain should proceed, e.g., on the 100  $\mu$ s time scale, then two initially different (and at L still present) cycles could subsequently merge. However, independent of this detail, one should remark that during the past decades the BR community has been quite reluctant to accept the additional complication of a conformational heterogeneity for BR (see the review by Lanyi<sup>3</sup> for a discussion). However, the results of the HRE simulations forced us to give up our corresponding resistance.

As is the case in any MD approach, the predictive power of our study remains limited by force field uncertainties and by the sampling problem. To gain an improved predictive power, we have done our best to partially master these problems specifically for the given case of BR.

Addressing the force field problem, we have calculated by DFT/MM the polarized force field PBR for the reference structure 1C3W/MD. Now one could suspect that this force field preferentially stabilizes BR in the crystallographic conformation  $C_1$ . To check this issue, we have additionally calculated another polarized force field PBR( $C_2$ ) for a reference structure representative for conformation  $C_2$ . This structure was obtained as a snapshot from a PBR trajectory initialized and staying in  $C_2$ . As shown in the SI, with PBR( $C_2$ ), the 300 K free energy

landscape  $\Delta G(d_1, d_2)$  of conformation  $C_2$  is nearly identical to that obtained with PBR.

Addressing the sampling problem, we have designed a HRE procedure for the lysine conformations. This procedure actually solved a part of the problem, because it proved the existence and important properties of alternative conformations. However, it also fell short of solving it completely because we had chosen the region of “hot” sampling a little too small.

### Summary and Outlook

Driven by our desire to construct a MD simulation system, which allows us to sample the room-temperature fluctuations within the chromophore binding pocket of BR in a reliable fashion, we have (i) derived the structure-adapted polarized force field PBR from self-consistent “moving domain” DFT/MM computations of partial charge distributions and have (ii) designed and applied a Hamiltonian replica exchange approach for probing the equilibrium fluctuations of the lysine chain, which is the covalent linkage between the RSBH<sup>+</sup> and the protein backbone.

For physiological temperatures, our HRE approach has revealed the coexistence of at least two major conformations  $C_1$  and  $C_2$ , whose contributions to the conformational ensemble show a distinct and opposite temperature dependence. Here, the enthalpically favored (crystallographic<sup>40</sup>) conformation  $C_1$  dominates at low temperatures, whereas the entropically favored conformation  $C_2$  (featuring a more flexible lysine chain) dominates at elevated temperatures and in the NMR ensemble.<sup>48</sup> The enhanced flexibility of  $C_2$  is due to a local melting of helix G at LyR216, which is enabled by a preformed kink (“ $\pi$ -bulge”) within that helix at Ala215.

According to the room-temperature PBR-MD simulations, the conformations  $C_1$  and  $C_2$  both preserve the stable H-bonded network generated by the small water cluster within the chromophore binding pocket. Surprisingly, a DFT/MM treatment has demonstrated that the dipole moments of the H<sub>2</sub>O molecules within that cluster are larger than in a bulk liquid, thus indicating stiffer binding interactions. The PBR-MD simulations have also shown that at room temperature transitions between  $C_1$  and  $C_2$  occur on time scales above 0.1  $\mu$ s. The thus emerging conformational heterogeneity of BR supports previous (and persistently doubted) interpretations of time-resolved resonance Raman spectra.<sup>92</sup>

The success of generating a reasonably stable MD model of BR now has paved the way toward the DFT/MM computation of the BR vibrational spectra, which will be tackled in part II of this work (DOI 10.1021/jp902432e). Why this task is still a highly interesting challenge will be discussed in the introductory remarks to part II (DOI 10.1021/jp902432e).

**Acknowledgment.** The authors gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (SFB 533, projects C1 and C3) and valuable discussions with B. Schropp concerning the properties of water.

**Supporting Information Available:** On 14 pages, the SI provides 8 tables, 10 figures, and text explaining and discussing the additional material. In the above text, the contents of the SI have been addressed at various locations. The SI specifies the force field PBR (tables 2–9), explains the classification of the lysine conformations (including the quoted NMR data, pp S5–S8), and presents and discusses results of MD simulations for the three conformations  $C_1$ – $C_3$  and for different force fields (pp. S9–S13). This material is available free of charge via the Internet at <http://pubs.acs.org>.

### References and Notes

- Oesterhelt, D.; Stoeckenius, W. *Nat. New Biol.* **1971**, 233, 149–152.
- Oesterhelt, D.; Stoeckenius, W. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, 70, 2853–2857.
- Lanyi, J. K. *Annu. Rev. Physiol.* **2004**, 66, 665–688.
- Haupts, U.; Tittor, J.; Oesterhelt, D. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, 28, 367–399.
- Garczarek, F.; Gerwert, K. *Nature* **2006**, 439, 109–112.
- McCamant, D.; Kukura, P.; Mathies, R. *J. Phys. Chem. B* **2005**, 105, 10449–10457.
- Herbst, J.; Heyne, K.; Diller, R. *Science* **2002**, 297, 822–825.
- Siebert, F.; Hildebrandt, P. *Vibrational spectroscopy in life science*; Wiley-VCH: Weinheim, Germany, 2007.
- Smith, S. O.; Myers, A.; Pardo, J.; Winkel, C.; Mulder, P.; Lugtenburg, J.; Mathies, R. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, 81, 2055–2059.
- Alshuth, T.; Stockburger, M. *Photochem. Photobiol.* **1986**, 43, 55–66.
- Fodor, S. P. A.; Pollard, W. T.; Gebhard, R.; van den Berg, E. M. M.; Lugtenburg, J.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85, 2156–2160.
- Siebert, F. *Methods Enzymol.* **1995**, 246, 501–526.
- Engelhard, M.; Gerwert, K.; Hess, B.; Kreutz, W.; Siebert, F. *Biochemistry* **1985**, 24, 400–407.
- Zscherp, C.; Heberle, J. *J. Phys. Chem. B* **1997**, 101, 10542–10547.
- Curry, B.; Broek, A.; Lugtenburg, J.; Mathies, R. A. *J. Am. Chem. Soc.* **1982**, 104, 5274–5286.
- Smith, S. O.; Myers, A.; Mathies, R.; Pardo, J.; Winkel, C.; van den Berg, E.; Lugtenburg, J. *Biophys. J.* **1985**, 47, 653–664.
- Großjean, M. F.; Tavan, P.; Schulten, K. *Eur. Biophys. J.* **1989**, 16, 341–349.
- Fahmy, K.; Großjean, M. F.; Siebert, F.; Tavan, P. *J. Mol. Struct.* **1989**, 214, 257–288.
- Großjean, M. F.; Tavan, P.; Schulten, K. *J. Phys. Chem.* **1990**, 94, 8059–8069.
- Großjean, M. F.; Tavan, P. *J. Chem. Phys.* **1988**, 88, 4884–4896.
- Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136, B864–B870.
- Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, 140, A1133–A1138.
- Frisch, M. J.; et al. *Gaussian 92*, revision F.2; Gaussian, Inc.: Pittsburgh PA, 1993.
- Nonella, M.; Tavan, P. *Chem. Phys.* **1995**, 199, 19–32.
- Zhou, X.; Mole, S. J.; Liu, R. *Vib. Spectrosc.* **1996**, 12, 73–79.
- Neugebauer, J.; Hess, B. *J. Chem. Phys.* **2004**, 118, 12247–12258.
- Schmitz, M.; Tavan, P. On the art of computing the IR spectra of molecules in condensed phase. In *Modern methods for theoretical physical chemistry of biopolymers*; Starikov, E. B., Tanaka, S., Lewis, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2006.
- Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, 110, 10452–10467.
- Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, 107, 316–322.
- Nonella, M.; Mathias, G.; Tavan, P. *J. Phys. Chem. A* **2003**, 107, 8638–8647.
- Klähn, M.; Mathias, G.; Koetting, C.; Nonella, M.; Schlitter, J.; Gerwert, K.; Tavan, P. *J. Phys. Chem. A* **2004**, 108, 6186–6194.
- Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, 121, 12233–12246.
- Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, 121, 12247–12258.
- Schrader, T. E.; Schreier, W. J.; Cordes, T.; Koller, F. O.; Babitzki, G.; Denschlag, R.; Renner, C.; Dong, S.; Löweneck, M.; Moroder, L.; Tavan, P.; Zinth, W. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104, 15729–15734.
- Schropp, B.; Tavan, P. *J. Phys. Chem. B* **2008**, 112, 6233–6240.
- Schultheis, V.; Reichold, R.; Schropp, B.; Tavan, P. *J. Phys. Chem. B*, submitted for publication, 2008.
- Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, 103, 227–249.
- Klähn, M.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2005**, 88, 3829–3844.
- Eichinger, M. Berechnung molekularer Eigenschaften in komplexer Lösungsumgebung: Dichtefunktionaltheorie kombiniert mit einem Molekularmechanik-Kraftfeld. Doktorarbeit, Ludwig-Maximilians Universität München, Germany, 1999.
- Luecke, H.; Schobert, B.; Richter, H.-T.; Cartiailler, J.-P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, 291, 899–911.
- Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, 259, 393–421.
- Essen, L.-O.; Siegert, R.; Lehmann, W. D.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 11673–11678.
- Hildebrandt, P.; Stockburger, M. *Biochemistry* **1984**, 23, 5539–5548.
- de Groot, H. J. M.; Harbison, S.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **1989**, 28, 3346–3353.
- MacKerell, A.; et al. *J. Phys. Chem. B* **1998**, 102, 3586–3616.



- (46) Alshuth, T. Kinetische und strukturelle Untersuchungen am Chromophor von Bacteriorhodopsin mit Hilfe zeitaufgelöster Resonanz-Raman-Spektroskopie. Doktorarbeit, Goerg-August Universität, Göttingen, 1985.
- (47) Gerwert, K. Transduktion der Lichtenergie in Protonen-Transfer-Reaktionen beim Bacteriorhodopsin: Eine Untersuchung mit Hilfe der zeitaufgelösten IR- und statischen FTIR-Differenzspektroskopie. Doktorarbeit, Albert-Ludwigs-Universität Freiburg im Breisgau, 1985.
- (48) Patzelt, H.; Simon, B.; terLaak, A.; Kessler, B.; Kühne, R.; Schmieder, P.; Oesterheld, D.; Oshkinat, H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9765–9770.
- (49) Baudry, J.; Tajkhorshid, E.; Molnar, F.; Phillips, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 905–918.
- (50) Kandt, C.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2004**, *86*, 705–717.
- (51) Grudinin, S.; Büldt, G.; Gordeliy, V.; Baumgaertner, A. *Biophys. J.* **2005**, *88*, 3252–3261.
- (52) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (53) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (55) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281–1297.
- (56) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 453–462.
- (57) Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.
- (58) Tavan, P.; Carstens, H.; Mathias, G. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In *Protein Folding Handbook. Part 1*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: Weinheim, Germany, 2005.
- (59) Warshel, A.; Kato, M.; Pislakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (60) Brünger, A. "X-PLOR Manual", The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1992.
- (61) Braiman, M.; Mogi, T.; Marti, T.; Stern, L.; Khorana, H.; Rothschild, K. *Biochemistry* **1988**, *27*, 8516–8520.
- (62) Gerwert, K.; Hess, B.; Soppa, J.; Oesterheld, D. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4943–4947.
- (63) Metz, G.; Siebert, F.; Engelhard, M. *FEBS Lett.* **1992**, *303*, 237–241.
- (64) Spassov, V. Z.; Luecke, H.; Bashford, D.; Gerwert, K. *J. Mol. Biol.* **2001**, *312*, 203–219.
- (65) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670–681.
- (66) Tavan, P.; Schulten, K.; Oesterheld, D. *Biophys. J.* **1985**, *47*, 415–430.
- (67) Tajkhorshid, E.; Paizs, B.; Suhai, S. *J. Phys. Chem. B* **1999**, *103*, 4518–4527.
- (68) Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S. *Biophys. J.* **2000**, *78*, 683–693.
- (69) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847–10860.
- (70) Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994**, *101*, 734–748.
- (71) Niedermeier, C.; Tavan, P. *Mol. Simul.* **1996**, *17*, 57–66.
- (72) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729–1749.
- (73) Grubmüller, H.; Tavan, P. *J. Comput. Chem.* **1998**, *19*, 1534–1552.
- (74) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. *J. Comput. Chem.* **2001**, *22*, 501–508.
- (75) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (76) Lingenheil, M.; Denschlag, R.; Reichold, R.; Tavan, P. *J. Chem. Theory Comput.* **2008**, *4*, 1293–1306.
- (77) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175–186.
- (78) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parinello, M. "CPMD V3.9", Copyright IBM Corp and MPI für Festkörperforschung Stuttgart, 2004; www.cpmd.org.
- (79) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (80) Perdew, J.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- (81) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2005.
- (82) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (83) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (84) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (85) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.
- (86) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.
- (87) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (88) Carstens, H.; Renner, C.; Milbradt, A. G.; Moroder, L.; Tavan, P. *Biochemistry* **2005**, *44*, 4829–4840.
- (89) Kloppenburg, M.; Tavan, P. *Phys. Rev. E* **1997**, *55*, 2089–2092.
- (90) Albrecht, S.; Busch, J.; Kloppenburg, M.; Metze, F.; Tavan, P. *Neural Networks* **2000**, *13*, 1075–1093.
- (91) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2008**, *458*, 244–248.
- (92) Diller, R.; Stockburger, M. *Biochemistry* **1988**, *27*, 7641–7651.
- (93) Certain details of the conclusions of Diller and Stockburger<sup>92</sup> were subsequently revised in a time-resolved Resonance-Raman study<sup>96</sup> on the pH dependence of the BR photocycle intermediates following L without, however, affecting the main result on the heterogeneity of L.
- (94) Zimanyi, L.; Saltiel, J.; Brown, L. S.; Lanyi, J. K. *J. Phys. Chem. A* **2006**, *110*, 2318–2321.
- (95) Hendler, R. W.; Shrager, R. I.; Meuse, C. W. *Biochemistry* **2008**, *47*, 5406–5416.
- (96) Eisfeld, W.; Pusch, C.; Diller, R.; Lahrmann, R.; Stockburger, M. *Biochemistry* **1993**, *32*, 7196–7215.

JP902428X

Supplementary material to the manuscript

Polarization effects stabilize  
bacteriorhodopsin's chromophore binding pocket:  
A molecular dynamics study.

G. Babitzki, R. Denschlag, and P. Tavan\*

Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität,  
Oettingenstr. 67, 80538 München, Germany

\*corresponding author, email: [tavan@physik.uni-muenchen.de](mailto:tavan@physik.uni-muenchen.de),  
phone: +49-89-2180-9220, fax: +49-89-2180-9202

## The partial charges of the PBR force field

The following tables 2-9 document the ESP partial charges  $Q_i$  (see Ref. 1 for definition) obtained by our self-consistent "moving domain DFT/MM approach"<sup>2,3</sup> for the atoms  $i$  of the BR chromophore LyR216 (Tab. 2) and for 22 nearby residues (Tab. 3-9). The atom names employed for the usual protein residues follow the CHARMM22 convention,<sup>4</sup> the atom names used for the special residue LyR216 are explained by Fig. 13. CHARMM22 partial charges<sup>4</sup> are designated as  $Q_{C22}$ .

Table 2: ESP partial charges for the RSBH<sup>+</sup> in BR from DFT/MM

atom	$Q_{LyR216}$	atom	$Q_{LyR216}$	atom	$Q_{LyR216}$	atom	$Q_{LyR216}$
C	0.56	HE2	0.08	H10	0.13	H4A	0.00
O	-0.73	NZ	-0.25	C9	0.10	H4B	0.00
N	-0.56	HZ1	0.40	C19	-0.06	C3	-0.05
HN	0.39	C15	0.07	H19A	0.04	H3A	0.01
CA	0.33	H15	0.13	H19B	0.04	H3B	0.01
HA	0.01	C14	-0.27	H19C	0.04	C2	-0.16
CB	-0.23	H14	0.18	C8	-0.24	H2A	0.03
HB1	0.05	C13	0.20	H8	0.16	H2B	0.03
HB2	0.05	C20	-0.15	C7	-0.01	C1	0.56
CG	0.24	H20A	0.07	H7	0.12	C16	-0.15
HG1	0.00	H20B	0.07	C6	-0.31	H16A	0.00
HG2	0.00	H20C	0.07	C5	0.06	H16B	0.00
CD	-0.13	C12	-0.19	C18	-0.14	H16C	0.00
HD1	0.05	H12	0.12	H18A	0.05	C17	-0.27
HD2	0.05	C11	-0.02	H18B	0.05	H17A	0.05
CE	0.01	H11	0.14	H18C	0.05	H17B	0.05
HE1	0.08	C10	-0.19	C4	0.13	H17C	0.05

Table 3: Partial charges for Ala53, Ala215, and Ser141

atom	$Q_{Ala53}$	$Q_{Ala215}$	$Q_{C22}$	atom	$Q_{Ser141}$	$Q_{C22}$
CB	-0.33	-0.33	-0.27	CB	0.23	0.05
HB1	0.11	0.11	0.09	HB1	0.04	0.09
HB2	0.11	0.11	0.09	HB2	0.04	0.09
HB3	0.11	0.11	0.09	OG	-0.77	-0.66
				HG1	0.46	0.43

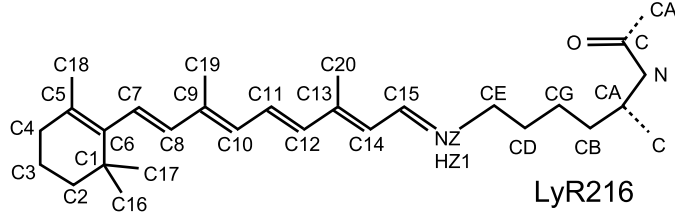


Figure 13: CHARMM-type names used in Tab. 2 for the heavy atoms within the special residue LyR216. In Tab. 2 the names of the hydrogen atoms point to the names of the attached heavy atoms. Thus, for instance, the methyl H-atoms at carbon C20 are called H20A, H20B, and H20C.

Table 4: Partial charges for Asp85, Asp212 and Arg82

atom	$Q_{\text{Asp85}}$	$Q_{\text{Asp212}}$	$Q_{\text{C22}}$	atom	$Q_{\text{Arg82}}$	$Q_{\text{C22}}$	atom	$Q_{\text{Arg82}}$	$Q_{\text{C22}}$
C	0.67	0.57	0.51	C	0.70	0.51	CD	0.11	0.20
O	-0.68	-0.75	-0.51	O	-0.70	-0.51	HD1	0.06	0.09
N	-0.69	-0.65	-0.47	N	-0.72	-0.47	HD2	0.06	0.09
HN	0.35	0.40	0.31	HN	0.36	0.31	NE	-0.62	-0.70
CA	0.35	0.43	0.07	CA	0.36	0.07	HE	0.40	0.44
HA	0.00	0.00	0.09	HA	0.00	0.09	CZ	0.87	0.64
CB	-0.53	-0.36	-0.28	CB	-0.33	-0.18	NH1	-0.87	-0.80
HB1	0.18	0.05	0.09	HB1	0.09	0.09	HH11	0.45	0.46
HB2	0.18	0.05	0.09	HB2	0.09	0.09	HH12	0.45	0.46
CG	1.01	1.16	0.62	CG	0.16	-0.18	NH2	-1.00	-0.80
OD1	-0.96	-0.87	-0.76	HG1	0.07	0.09	HH21	0.47	0.46
OD2	-0.88	-1.03	-0.76	HG2	0.07	0.09	HH22	0.47	0.46

Table 5: Partial charges for Thr89, Thr90, and Thr142

atom	$Q_{\text{Thr89}}$	$Q_{\text{Thr90}}$	$Q_{\text{Thr142}}$	$Q_{\text{C22}}$
CB	0.35	0.26	0.34	0.14
HB	0.00	0.00	0.00	0.09
OG1	-0.81	-0.80	-0.73	-0.66
HG1	0.55	0.50	0.50	0.43
CG2	-0.45	-0.17	-0.26	-0.27
HG21	0.12	0.07	0.05	0.09
HG22	0.12	0.07	0.05	0.09
HG23	0.12	0.07	0.05	0.09

Table 6: Partial charges for Met20, Met118, Met145, and Val49

atom	$Q_{\text{Met20}}$	$Q_{\text{Met118}}$	$Q_{\text{Met145}}$	$Q_{\text{C22}}$	atom	$Q_{\text{Val49}}$	$Q_{\text{C22}}$
CB	-0.19	-0.20	-0.13	-0.18	CB	0.15	-0.09
HB1	0.06	0.11	0.09	0.09	HB	0.01	0.09
HB2	0.06	0.11	0.09	0.09	CG1	-0.19	-0.27
CG	0.24	0.18	0.12	-0.14	HG11	0.02	0.09
HG1	0.08	0.02	0.00	0.09	HG12	0.02	0.09
HG2	0.08	0.02	0.00	0.09	HG13	0.02	0.09
SD	-0.48	-0.40	-0.29	-0.09	CG2	-0.21	-0.27
CE	0.03	0.01	-0.12	-0.22	HG21	0.06	0.09
HE1	0.04	0.05	0.08	0.09	HG22	0.06	0.09
HE2	0.04	0.05	0.08	0.09	HG23	0.06	0.09
HE3	0.04	0.05	0.08	0.09			

Table 7: Partial charges for Pro186 and Leu93

atom	$Q_{\text{Pro186}}$	$Q_{\text{C22}}$	atom	$Q_{\text{Leu93}}$	$Q_{\text{C22}}$
C	0.50	0.51	CB	-0.40	-0.18
O	-0.64	-0.51	HB1	0.08	0.09
N	-0.33	-0.29	HB2	0.08	0.09
CA	0.30	0.02	CG	0.71	-0.09
HA	0.00	0.09	HG	-0.06	0.09
CD	-0.09	0.00	CD1	-0.20	-0.27
HD1	0.10	0.09	HD11	0.01	0.09
HD2	0.10	0.09	HD12	0.01	0.09
CB	-0.32	-0.18	HD13	0.01	0.09
HB1	0.07	0.09	CD2	-0.42	-0.27
HB2	0.07	0.09	HD21	0.06	0.09
CG	0.20	-0.18	HD22	0.06	0.09
HG1	0.02	0.09	HD23	0.06	0.09
HG2	0.02	0.09			

Table 8: Partial charges for Trp86, Trp138, Trp182, and Trp189

atom	$Q_{\text{Trp86}}$	$Q_{\text{Trp138}}$	$Q_{\text{Trp182}}$	$Q_{\text{Trp189}}$	$Q_{\text{C22}}$
CB	-0.10	-0.13	0.06	0.01	-0.180
HB1	0.06	0.12	0.01	0.07	0.090
HB2	0.06	0.12	0.01	0.07	0.090
CG	-0.02	-0.16	-0.19	-0.20	-0.030
CD2	0.13	0.04	0.04	0.08	-0.020
CD1	-0.01	0.02	0.09	-0.08	0.035
HD1	0.17	0.12	0.14	0.15	0.115
NE1	-0.60	-0.65	-0.67	-0.36	-0.610
HE1	0.44	0.43	0.46	0.36	0.380
CE2	0.25	0.37	0.39	0.17	0.130
CE3	-0.21	-0.19	-0.19	-0.22	-0.115
HE3	0.10	0.14	0.14	0.10	0.115
CZ2	-0.25	-0.31	-0.35	-0.24	-0.115
HZ2	0.13	0.14	0.13	0.14	0.115
CZ3	-0.18	-0.20	-0.17	-0.15	-0.115
HZ3	0.07	0.10	0.09	0.08	0.115
CH2	-0.13	-0.04	-0.06	-0.08	-0.115
HH2	0.09	0.08	0.07	0.10	0.115

Table 9: Partial charges for Tyr57, Tyr83, and Tyr185

atom	$Q_{\text{Tyr57}}$	$Q_{\text{Tyr83}}$	$Q_{\text{Tyr185}}$	$Q_{\text{C22}}$
CB	-0.22	-0.01	-0.23	-0.180
HB1	0.10	0.00	0.05	0.090
HB2	0.10	0.00	0.05	0.090
CG	0.15	0.11	0.14	0.000
CD1	-0.09	-0.16	-0.14	-0.115
HD1	0.17	0.07	0.17	0.115
CD2	-0.27	-0.12	-0.19	-0.115
HD2	0.11	0.17	0.10	0.115
CE1	-0.35	-0.21	-0.22	-0.115
HE1	0.18	0.13	0.12	0.115
CE2	-0.23	-0.29	-0.26	-0.115
HE2	0.15	0.10	0.16	0.115
CZ	0.45	0.51	0.44	0.110
OH	-0.79	-0.84	-0.73	-0.540
HH	0.54	0.54	0.54	0.430

## Analysis of lysine conformations in the HRE ensemble

As explained in Methods a maximum-likelihood density estimate  $\rho(\Psi)$  was generated for the 300 K HRE trajectory  $\Psi(t) \in [0, 2\pi]^5$  of torsional angles within the lysine chain by optimizing the parameters of a  $M = 100$  component normal mixture. The  $M$  univariate normal distributions contribute with equal weights  $P_r = 1/M$ ,  $r = 1, \dots, M$  to the mixture and have identical widths  $\sigma^{\text{ML}}$ . The mixture is supposed to model the invariant (i.e. equilibrium) density associated with the 300 K trajectory. Smoothing of  $\rho(\Psi)$  at decreasing resolutions by convolution with a Gaussian of width  $\sigma$  yields a scale space of model densities  $\rho(\Psi|\sigma)$ , whose  $N_\sigma$  maxima point to prototypical conformations  $\psi_\beta(\sigma)$ ,  $\beta = 1, \dots, N_\sigma$ .

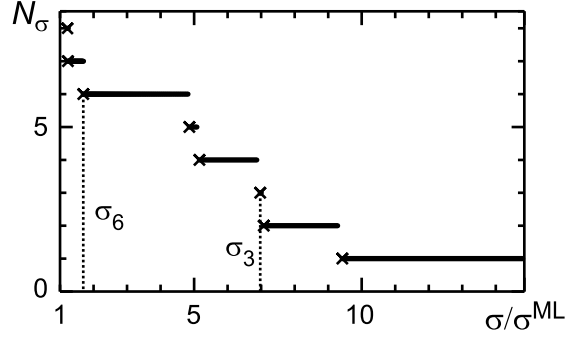


Figure 14: Number  $N_\sigma$  of maxima of the smoothed maximum likelihood density estimate  $\rho(\Psi|\sigma)$  for the 300 K distribution of lysine chain torsion angles in the cyclic space  $[0, 2\pi]^5$  as a function of the resolution  $\sigma$ . Crosses mark lower bounds of  $\sigma$  ranges associated to a given  $N_\sigma$ . For  $N_\sigma = 3$  and  $N_\sigma = 6$  these lower bounds are marked by dashed lines.

Fig. 14 shows the number  $N_\sigma$  of maxima of  $\rho(\Psi|\sigma)$  as a function of  $\sigma$ .  $N_\sigma$  first slowly increases with decreasing  $\sigma$  until it switches to a rapid increase at  $\sigma < \sigma_6$ . For the classification of the lysine chain conformations we have chosen the coarse resolution  $\sigma_3$  featuring  $N_\sigma = 3$  conformational states. This choice was motivated by the finding that all finer resolutions ( $N_\sigma > 3$ ) predict conformational states with a nearly vanishing ( $< 0.1\%$ ) occurrence in the 300 K ensemble. These outliers states can safely be neglected for the description of this ensemble. Nevertheless we will additionally present a more highly resolved conformational analysis using  $N_\sigma = 6$  states further below.

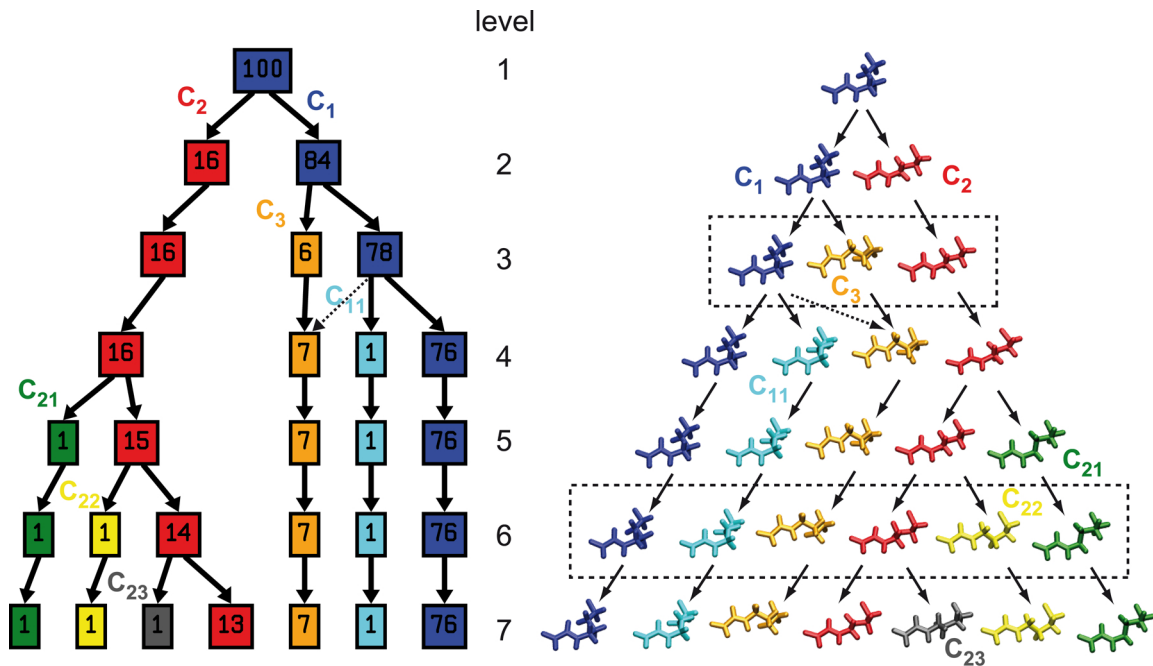


Figure 15: (a) Hierarchical decomposition of the 100 component mixture density  $\rho(\Psi)$  into class local densities  $\rho_\beta(\Psi|l)$ ,  $\beta = 1, \dots, l$ , with  $M_{l\beta}$  components at level  $l$  and (b) associated hierarchy of prototypical conformations (the two trees are drawn as mirror images). In (a) the numbers  $M_{l\beta}$  in the square boxes count the univariate normal distributions contributing to the respective local model  $\rho_\beta(\Psi|l)$ . Thus, the  $M_{l\beta}$  indicate at each level  $l$  the statistical weights of the class local densities  $\rho_\beta(\Psi|l)$  within the total mixture.

The resulting hierarchical decomposition of the mixture density  $\rho(\Psi)$  into class local densities  $\rho_\beta(\Psi|l)$ ,  $\beta = 1, \dots, l \equiv N_\sigma$ , is depicted in Fig. 15a. The local densities  $\rho_\beta(\Psi|l)$  have  $M_{l\beta}$  components ( $\sum_{\beta=1}^l M_{l\beta} = M$ ) and a single maximum after convolution with a Gaussian of width  $\sigma_l$ . Fig. 15b translates the combination  $\Psi_{l\beta}$  of torsional angles marking the respective maximum back into a lysine chain structure, i.e., into a prototypical conformation. In the paper the prototypes at level  $l = 3$  were used for the classification of the trajectories (Figs. 4, 11, 12).

When applying the prototypes at level  $l = 6$  to the classification of the HRE trajectories one obtains the results summarized by Fig. 16. Here, Fig. 16a is the higher resolution analogue to Fig. 11 and Fig. 16b that of Fig. 12. Comparing first Fig. 16b with Fig. 12 one immediately sees that the inclusion of additional conformational states does not change the description of the 300 K ensemble. During all the shown 5 ns time spans of HRE simulation the additional conformations hardly ever were detected in the 300 K replica.

A slightly different picture emerges when one considers the contribution of the additional states to the higher temperature replicas during the last 25 ns of HRE simulation. According to Fig. 16a one of the off-springs of the red conformation  $C_2$ , i.e., the yellow conformation  $C_{22}$  is found at sizable frequencies at higher temperatures. In Fig. 11 this conformation had been still identified with its parent  $C_2$ .

In summary, for the HRE simulations the higher resolution classification simply confirms the low resolution results. Similarly one does not find any changes when using the six state model



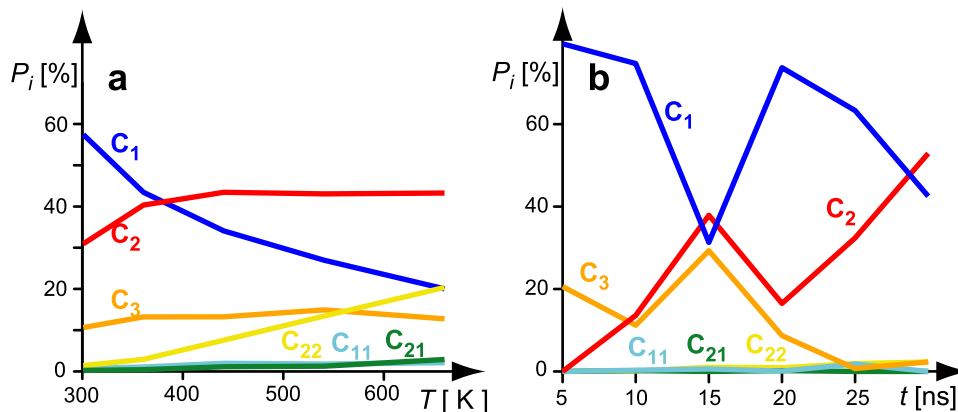


Figure 16: (a) Occupancies of the conformational states at level 6 during the last 25 ns of the HRE simulation as a function of the replica temperature. (b) Temporally resolved occupancies at 300 K; shown are averages  $\langle P_i \rangle_{\Delta t}(t_k)$  over time spans  $\Delta t = 5$  ns.

instead of the three state model for the analysis of the various 300 K MD trajectories (data not shown).

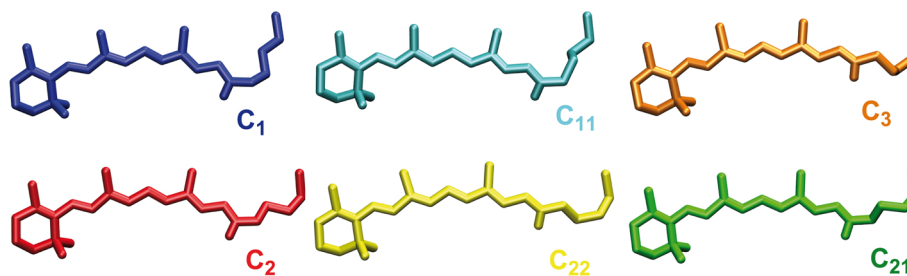


Figure 17: Prototypical structures of LyR216 identified at level  $l = 6$ .

However, slight differences between the three- and six-state models appear in the classification of the twelve NMR structures.<sup>5</sup> Fig. 17 displays the prototypical structures of the LyR216 side chain identified by the six-state classifier. These structures should be visually compared with the corresponding results of the three-state classifier shown in Fig. 10. This comparison then clearly demonstrates that the prototypical structures identified by the classifier for the conformations  $C_1$ ,  $C_2$ , and  $C_3$  at the two levels of resolution are nearly identical and that the off-springs  $C_{11}$ ,  $C_{21}$ , and  $C_{22}$  exhibit distinct structural differences from the prototypes of their respective parent classes.

The result of the classification of the NMR structures is depicted in Fig. 18. One now finds that the X-ray conformation  $C_1$  is absent in the NMR ensemble and that the three structures previously assigned to  $C_1$  are in fact elements of the (light-blue) subclass  $C_{11}$ , which exhibits slight structural deviations from its parent class  $C_1$ . In the 300 K MD ensemble this subclass is quite rare (cf. Fig. 16b). Similarly, only four structures are still classified into  $C_2$ , whereas the remaining five structures are sorted into associated off-spring classes (four into the green class  $C_{21}$ , and one into the yellow class  $C_{22}$ ). Note in this connection that the yellow class  $C_{22}$

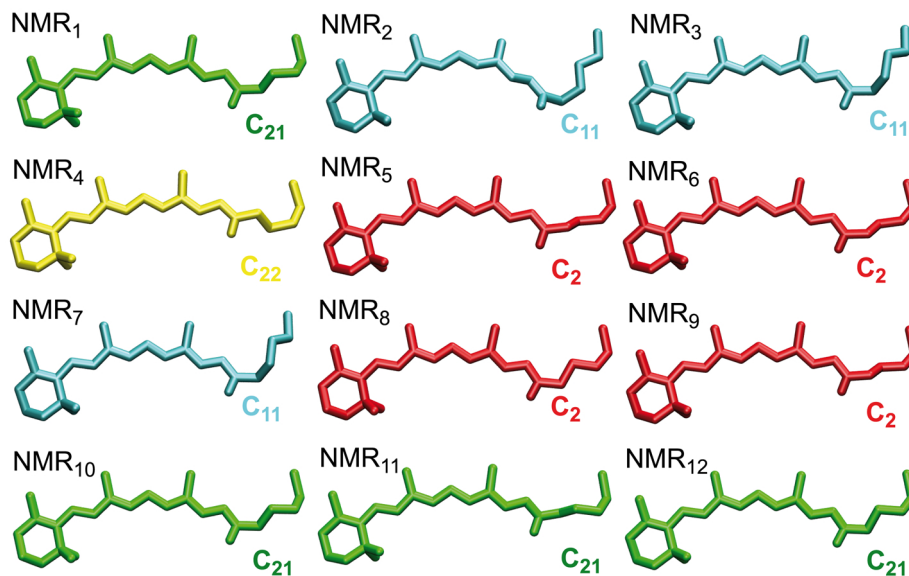


Figure 18: The twelve NMR structures of the LyR216 side chain as determined by Patzelt et al.<sup>5</sup> are classified by the six-state model. Classes are indicated by color coding.

occurs at sizable and the green class at enhanced frequencies in the high-temperature replicas (Fig. 16a). As remarked in the text of the paper the occurrence of such "high-temperature" structures in the NMR ensemble is no surprise, because the latter has been generated by MD based simulated annealing.

In this connection it is important to note that Patzelt et al.<sup>5</sup> used a different force field [AMBER, Ref. 6] in their annealing computations and imposed the constraints on the intra-helical H-bond distances **after** annealing. Because according to our simulation data, the shorter lysine conformations (e.g. C<sub>22</sub>) are strictly correlated with a partial melting of the helix G structure near LyR216, such a melting presumably occurred during the (high temperature) annealing refinement and produced such shortened conformations (independently of the force field). The subsequent restoration of the helix structure by restraining the intra-helical H-bond distances to the X-ray values then most likely left high-temperature configurations of the lysine chain frozen out.

## Trajectories and energy landscapes from 300 K MD

Figure 19 compares free energy landscapes, which were obtained from extended 300 K MD simulations for the three lysine conformations C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> and for different force fields. The top row (a-c) pertains to the polarized force field PBR, the bottom row to CHARMM. Because the crystallographic conformation C<sub>1</sub> is extremely stable in the PBR setting, the simulations for the other conformations had to be started with a respective structure. The landscapes in the first column (a,d) were already presented in Figs. 8 and 6.

According to Figs. 19c and 19f the conformation C<sub>3</sub> can stabilize the H-bond between the RSBH<sup>+</sup> and a water molecule independently of the force field. For the more frequent conformations C<sub>2</sub> and C<sub>1</sub>, however, this H-bond decays with CHARMM22 (Figs. 19d and 19e) and

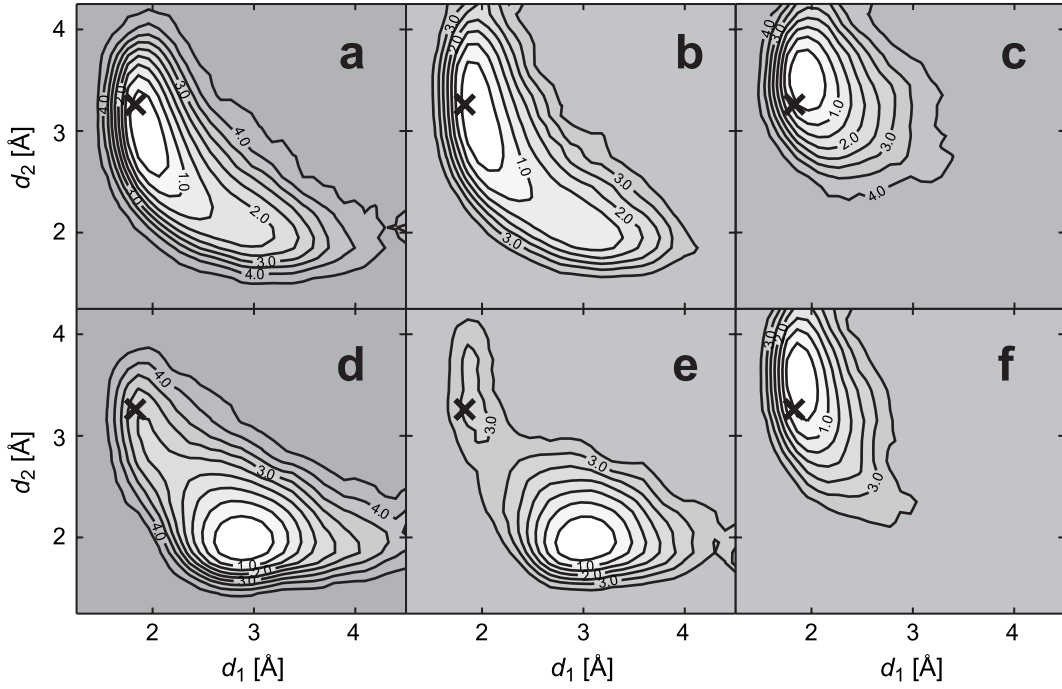


Figure 19: Free energy landscapes  $\Delta G(d_1, d_2)$  obtained from 300 K MD simulations of the three different lysine conformations  $C_1$  (a,d),  $C_2$  (b,e), and  $C_3$  (c,f). a-c: PBR, d-f: CHARMM. MD simulations lasted 50 ns (a,d), 20 ns (b,c), 13 ns (e), and 12 ns (f).

becomes stable solely upon the inclusion of polarization effects (Figs. 19a and 19b).

The 300 K MD trajectories  $\delta_\alpha(t)$ , from which Figs. 19a and 19d-e derive, are shown in Figs. 9 and 4, respectively. The trajectories underlying Figs. 19b and 19c were taken from 20 ns PBR-MD simulations of conformations  $C_2$  and  $C_3$ . These trajectories are shown in Fig. 20.

Apart from slightly larger fluctuations the PBR trajectories  $\delta_1$ - $\delta_5$  of conformation  $C_2$  in Fig. 20a closely resemble the corresponding trajectories of conformation  $C_1$  in Fig. 9. Marked differences exist for  $\delta_6(t)$ . Instead of the value of about 2.1 Å observed for  $C_1$  we now find about 3.3 Å. This is just the average value, which is assumed by  $\delta_6(t)$  in the CHARMM22 trajectory Fig. 4 whenever the system was in state  $C_2$  and indicates a loosening of helix G at LyR216.

The PBR-MD trajectories of conformation  $C_3$  in Fig. 20a are quite boring. Through the persistently small value  $\delta_2 \approx 2$  Å they indicate the H-bond between the RSBH<sup>+</sup> and one of the water molecules. In agreement with the sharp and deep free energy minimum displayed by Fig. 19c the fluctuations are very small pointing to a rock-solid arrangement. Furthermore, for  $C_3$  the distance  $\delta_6 \approx 4$  Å is larger than for  $C_2$  (compare Figs. 20a and 20b).

As is apparent from a comparison with the CHARMM trajectories in Fig. 4, the value of  $\delta_6$  is independent of the force field and is characteristic for the respective lysine conformation  $C_i$ . This result corroborates the validity of our conformational analysis, which had been exclusively based on a clustering of the dihedral angles within the lysine chain and did not consider other conformational coordinates like  $\delta_6$ . Furthermore it underlines the validity of our claim on the close connection between the loosening of Helix G and the appearance of alternative conformations.

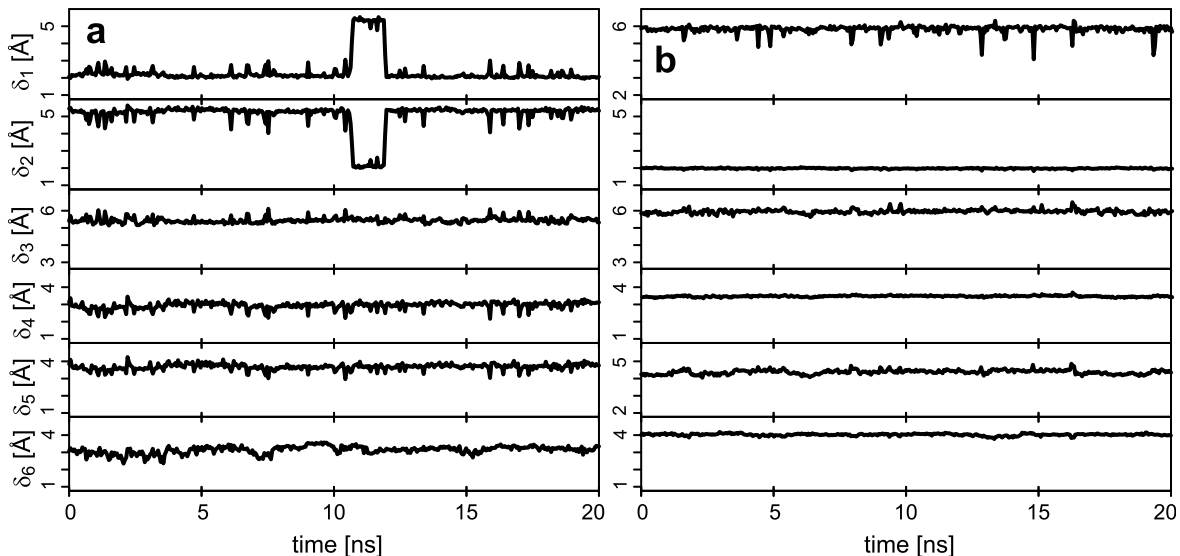


Figure 20: Trajectories of six distances  $\delta_\alpha(t)$ , from 20 ns PBR-MD simulations at 300 K starting at lysine conformations  $C_2$  (a) and  $C_3$  (b), respectively. For further explanations see the caption to Fig. 4.

## Energy landscapes for variant force fields

As mentioned in the Discussion we have carried out additional 300 K MD simulations for variant force fields, which were calculated by DFT/MM for different reference structures. With these simulations we wanted to find out, to what extent the free energy landscape depends on the details of the structure chosen for the DFT/MM computation of the polarized force field. One of the structures was a snapshot from the PBR trajectory of conformation  $C_2$  depicted in Fig. 20a and yielded a force field PBR( $C_2$ ).

Figs. 21a and 21c have been already presented above as parts a and b of the large Fig. 19. Thus, Fig. 21a repeats the free energy landscape computed for conformation  $C_1$  with the original force field PBR (which is tuned to this conformation). The landscape is compared with the one obtained for  $C_1$  with the field PBR( $C_2$ ), which is tuned to the other conformation  $C_2$  (Fig. 21b). The converse comparison is presented by Figs. 21c and 21d for conformation  $C_2$ . Here d pertains to the conformation local force field PBR( $C_2$ ), whereas c has been obtained with the force field PBR tuned to  $C_1$ . Quite apparently for each of the two conformations the landscapes obtained with the two differently polarized force fields are extremely similar. In fact, within statistical error they are indistinguishable. All landscapes indicate a persistently stable H-bond between the RSBH<sup>+</sup> and one of the water molecules. As a result we conclude that polarized force fields calculated for different structures featuring this H-bond can stabilize this H-bond in 300 K simulations.

We even checked whether we can compute a polarized force field performing as badly as CHARMM22. For this purpose we took the stable CHARMM22 structure, which is depicted in Fig. 5 and features a hydrogen bond between the RSBH<sup>+</sup> and Asp85, as the reference and computed by DFT/MM a corresponding polarized force field called PBR(CHARMM22).

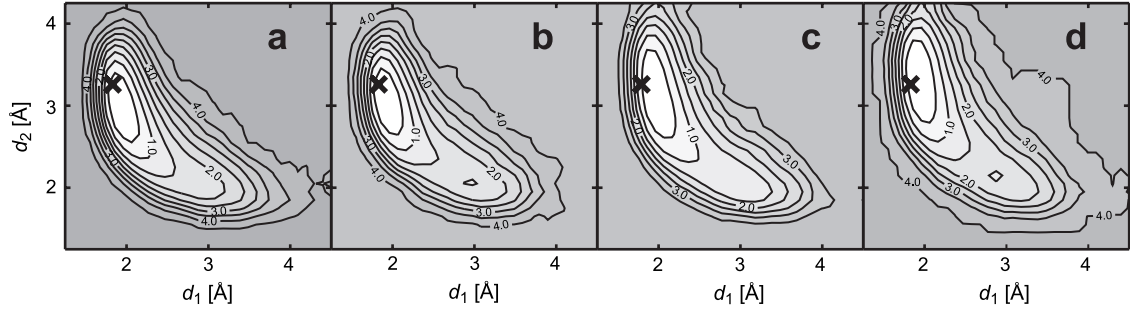


Figure 21: Free energy landscapes  $\Delta G(d_1, d_2)$  obtained from 20 ns, 300 K MD simulations of lysine conformations  $C_1$  (a,b) and  $C_2$  (c,d) for two different force fields. a,c: PBR as in Figs. 19a,b. b,d: PBR( $C_2$ )

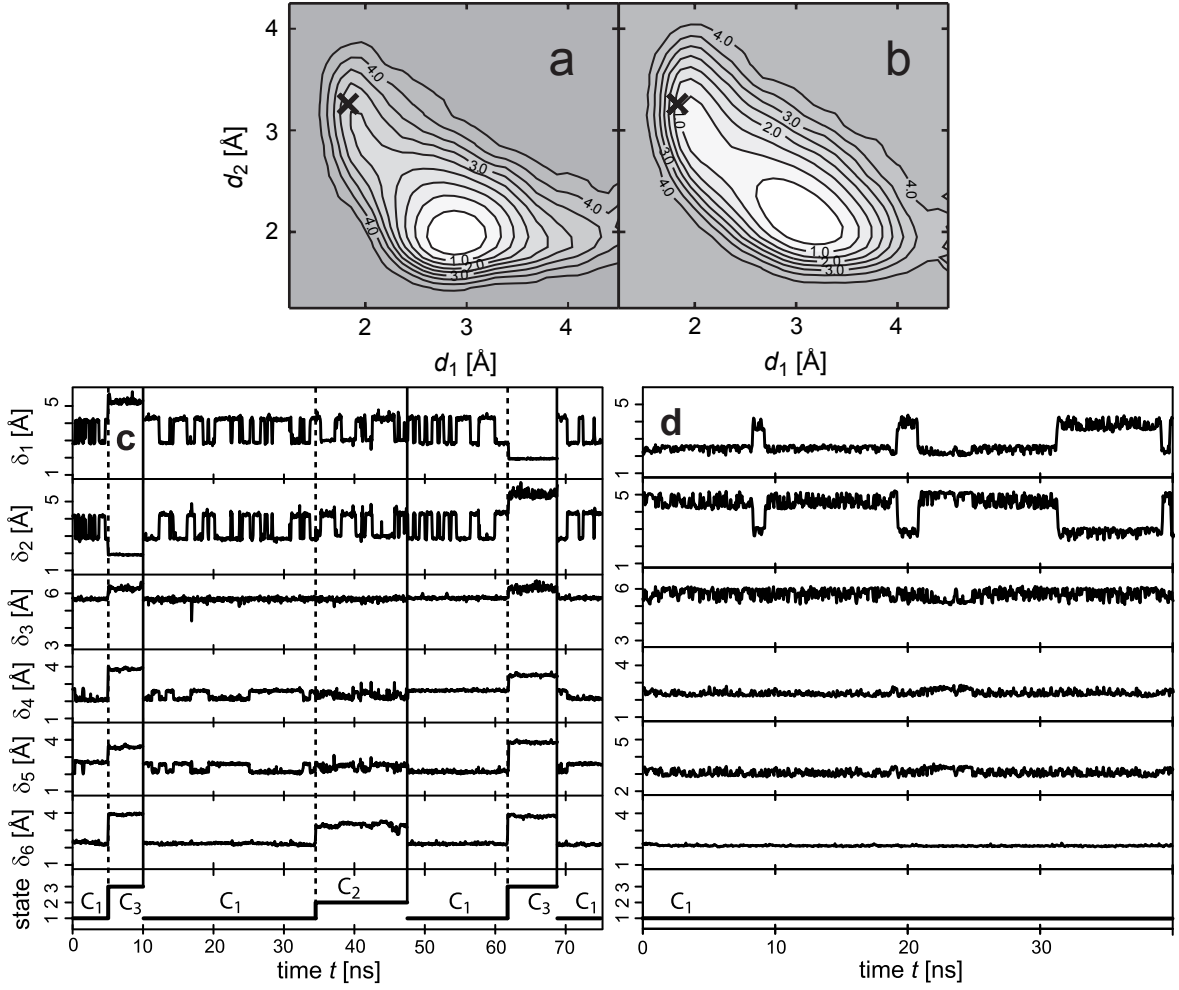


Figure 22: Results of 300 K MD simulations initialized at conformation  $C_1$  for two different force fields. (a-b): Free energy landscapes  $\Delta G(d_1, d_2)$  of  $C_1$ . a: 50 ns CHARMM22 (cf. Figs. 6 and 19d). b: 40 ns PBR(CHARMM22). (c-d): Trajectories. c: 75 ns CHARMM22 (cf. Fig. 4). d: 40 ns PBR(CHARMM22).

As is apparent from an inspection of Figs. 22a and 22b the free energy landscape  $\Delta G(d_1, d_2)$  of conformation  $C_1$  obtained with PBR(CHARMM22) looks quite similar to the one resulting for CHARMM22 (and previously shown in Figs. 6 and 19d). However, a closer comparison of Figs. 22a and 22b demonstrates that the polarized force field PBR(CHARMM22) cannot stabilize the wrongly assumed H-bond bond between the RSBH<sup>+</sup> as rigidly as CHARMM22. For PBR(CHARMM22) one sees larger fluctuations of the distance  $d_2$  between the Schiff base proton and an Asp85 oxygen. Furthermore it seems that fluctuations towards the correct H-bonded structure are somewhat more frequent than with CHARMM22.

The trajectory comparison in Figs. 22c and 22d confirms the existence of enhanced distance fluctuations within the hydrogen bonded network for PBR(CHARMM22). Concurrently it reveals, that the position exchange of the two water molecules  $W_1$  and  $W_2$  is much less frequent than with CHARMM22 thus indicating a somewhat more rigid binding of the closer water molecule to the Schiff base. But most importantly the 40 ns PBR(CHARMM22) trajectory shows

---

not any sign of a conformational transition to  $C_2$  or  $C_3$ , which in the CHARMM22 case occurs about every 20 ns. PBR(CHARMM22) apparently stabilizes  $C_1$  better than CHARMM22. In summary the PBR(CHARMM22) experiment demonstrates that even a completely ill-tuned polarized force field still performs better on the BR structure than CHARMM22. On the other hand this experiment does not say much more, because logics dictates that results derived from a false assumption may be equally well true or false.

## References

1. Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129-145.
2. Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175-186.
3. Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452-10467.
4. MacKerell, A. *et al.* *J. Phys. Chem. B* **1998**, *102*, 3586-3616.
5. Patzelt, H.; Simon, B.; terLaak, A.; Kessler, B.; Kühne, R.; Schmieder, P.; Oesterhelt, D.; Oschkinat, H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9765-9770.
6. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; K. M. Merz, J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.



## 2.2 DFT/MM-Schwingungsanalyse des Retinalchromophors von BR

Im nachfolgend abgedruckten Manuskript

Galina Babitzki, Gerald Matthias and Paul Tavan, : „The infra-red spectra of the retinal chromophore in bacteriorhodopsin calculated by a DFT/MM approach“,

das im *Journal of Physical Chemistry* publiziert ist, wird mit DFT/MM Methoden systematisch untersucht, welchen Einfluss diverse Faktoren auf die Struktur und die IR-Spektren des BR Chromphors haben. Dazu gehören die Sterik und Elektrostatik der Bindungstasche, die Temperatur, die Konformation der Lysinkette etc. Es wird anhand detaillierter Vergleiche mit wohlbekannten spektroskopischen Daten gezeigt, dass jede Annäherung der Modellbildung an die physikalische Realität (z.B. physiologische statt tiefe Temperaturen) zu einer besseren Beschreibung führt, dass aber dennoch gewisse Defizite verbleiben, die wahrscheinlich dem eingesetzten DFT Funktional angelastet werden müssen. Wie gezeigt wird, ist die Qualität der Beschreibung aber gut genug um grob falsche Modelle der Proteinstruktur von korrekten Modellen allein anhand berechneter IR Spektren zu unterscheiden.

Zu dieser Veröffentlichung gehört Zusatzmaterial in Form ergänzender Tabellen, Graphiken und Texte, die im Anschluss an das Manuskript ebenfalls abgedruckt sind.



# The Infrared Spectra of the Retinal Chromophore in Bacteriorhodopsin Calculated by a DFT/MM Approach

G. Babitzki, G. Mathias, and P. Tavan\*

Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität Oettingenstr. 67, 80538 München, Germany

Received: March 18, 2009; Revised Manuscript Received: May 25, 2009

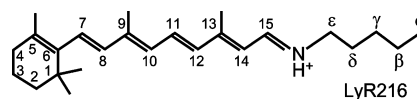
In the preceding paper (DOI 10.1021/jp902428x), we have derived the polarized force field “PBR” for bacteriorhodopsin (BR) using hybrid methods which combine density functional theory (DFT) with molecular mechanics (MM) models. This polarized force field has enabled extended molecular dynamics (MD) simulations of BR’s chromophore binding pocket which closely preserve the experimentally well-known structure. Here, we employ the PBR-MD trajectories obtained for the conformational substates prevalent at physiological temperatures as material for the DFT/MM computation of the chromophore’s vibrational spectra. By comparison with DFT results on the structure and vibrational spectra of an isolated chromophore, we identify the structural and spectral changes induced by the protein environment. Comparisons with the wealth of experimental data available in the literature on the chromophore’s vibrational spectra yield estimates on the accuracy of the DFT/MM descriptions. We discuss why highly accurate DFT/MM descriptions are expected to become a decisive tool for solving the long-standing enigma of how the light-driven proton pump BR actually works.

## Introduction

In the introductory remarks to the preceding paper,<sup>1</sup> which we will call “part I” of this work (DOI 10.1021/jp902428x), we have shortly reviewed the physiological role of the retinal protein bacteriorhodopsin (BR) (see ref 2 for a more extended review) and have indicated the reasons why this light-driven proton pump has served as a standard test system for first applications of new biophysical techniques. Clearly, one could perceive also our reported efforts<sup>1</sup> to generate a computer model of the chromophore binding pocket of BR remaining stable in quite extended molecular dynamics (MD) simulations as a purely methods oriented contribution. In view of the widespread jaundice that the mechanism of proton pumping in BR has been thoroughly clarified in the past, such a misperception would seem quite natural.

One of the sources of the quoted prejudice is a “News and Views” article that appeared nearly a decade ago in *Nature* under the heading “Bacteriorhodopsin—the movie”.<sup>3</sup> Here, we read in the abstract: “For 30 years and more, the mechanism of a microbial proton pump has been subject to increasingly sophisticated analysis. The full picture of how the pump operates is now emerging.” As we will argue below, there is no such “full picture” up to now and there are very good reasons to put the issue of the BR pump mechanism, to which one of the authors contributed a speculative suggestion three decades ago,<sup>4,5</sup> once again on the agenda. The corresponding analysis will then additionally answer the question why efforts to compute the vibrational spectra of the BR chromophore as accurately as possible are still of vital scientific interest.

This paper is organized as follows. First, we will shortly analyze why the tremendous efforts to clarify the proton pump mechanism in BR have not yet rendered a solution. Thereby, we will motivate our own efforts and, hopefully, inspire others. Subsequently, we will sketch the computational methods



**Figure 1.** In the light-adapted initial state BR<sub>568</sub> of the photocycle, the chromophore is the protonated Schiff base of all-*trans* retinal (all-*trans* RSBH<sup>+</sup>) covalently attached to the apoprotein through Lys216.

employed by us for the computation of the chromophore’s vibrational spectra. The corresponding results will be presented and discussed on the background of the available wealth of spectroscopic data.<sup>6–8</sup> A summary and an outlook conclude the paper.

**What is the Chromophore Structure in L?** Many experts in the field of BR are convinced that an answer to the above question would immediately lead to a deep and thorough understanding of the proton pump mechanism acting in BR (see, e.g., refs 4, 5, and 9–16). Here, “L” designates an early intermediate of the BR photocycle, which appears a few microseconds after the all-*trans* → 13-*cis* photoisomerization<sup>17</sup> of the BR<sub>568</sub> chromophore (see Figure 1 for the chromophore’s chemical structure in the light-adapted initial state BR<sub>568</sub>). The key role of the L intermediate for proton pumping immediately becomes apparent if one considers the fact that the first proton transfer during the proton pump cycle occurs in the L to M transition. Here, the proton originally attached to the RSBH<sup>+</sup> is transferred to a nearby aspartate (Asp85) and the key mechanistic questions are (i) why do the chromophore and/or Asp85 drastically change their pK<sub>a</sub>’s in the BR<sub>568</sub> to L transition and (ii) why is the subsequent proton transfer irreversible?<sup>4,5,18</sup>

Despite the fact that many X-ray structures<sup>16,19–22</sup> allegedly describing the L intermediate have been published, the detailed structure of this intermediate and particularly that of the chromophore have remained elusive up to now. The reason for this lack of knowledge is quite subtle:

All quoted attempts to characterize the L structure by crystallographic means were based on the same low-temperature

\* Corresponding author. E-mail: tavan@physik.uni-muenchen.de. Phone: +49-89-2180-9220. Fax: +49-89-2180-9202.

trapping technique. Similarly, also recent solid-state NMR data on the chromophore structure in L were obtained for a structure trapped at 150 K.<sup>23</sup> If the trapped low-temperature L state would represent a reasonable model for the functional L intermediate of the room-temperature proton pump cycle, then this approach would be valid. However, already some time ago Rödiger et al.<sup>14</sup> have shown by carefully comparing time-resolved Fourier transform infrared (FTIR) spectra with low-temperature ones that the structure of the chromophore is different in the two cases. The L intermediate occurring in the photocycle at physiological temperatures shows the so-called 15-Hoop band (where “15-Hoop” denotes the hydrogen out-of-plane vibration at C<sub>15</sub>) with a sizable intensity, indicating substantial twists around the C<sub>14</sub>–C<sub>15</sub> single bond of the RSBH<sup>+</sup>,<sup>24</sup> which finding has important implications concerning the pump mechanism.<sup>9–12</sup> In contrast, the corresponding low-temperature spectrum exhibits a vanishing intensity of the 15-Hoop indicating that, in the cryogenically trapped L state, the chromophore is essentially planar in the region of the C<sub>14</sub>–C<sub>15</sub> bond. Furthermore, in the L intermediate of the photocycle, the C<sub>14</sub>–C<sub>15</sub> stretching vibration is found at a different spectral location than in the trapped so-called L state.<sup>14</sup>

The spectra of Rödiger et al.<sup>14</sup> have recently been confirmed by Lórenz-Fonfría et al.<sup>25</sup> applying the same time-resolved step-scan FTIR technique to a wider spectral region (800–4000 cm<sup>−1</sup>). Concentrating on vibrational bands associated with internal water molecules, these authors additionally found that the distribution of internal water molecules in the L intermediate of the photocycle substantially differs from that in the trapped low-temperature L state.

These results show that low-temperature X-ray structures of trapped L can neither provide correct information on the chromophore structure nor provide correct information on the distribution of internal waters. Moreover, such X-ray structures cannot give information on the pumping mechanism because the trapped L state predominantly shows a shunt reaction back to BR<sub>568</sub> upon heating, as recently rediscovered by Dioumaev et al.<sup>26</sup> Thus, the low-temperature L, instead of proceeding to the M state through the deprotonation of the 13-*cis* RSBH<sup>+</sup>, i.e., instead of pumping protons, prefers to return directly to the initial state BR<sub>568</sub> through a thermal 13-*cis* to all-*trans* isomerization of the RSBH<sup>+</sup>.<sup>27</sup>

In summary, the trapped low-temperature “L” state does not represent a model for the L intermediate of the photocycle because it has a distinctly different IR spectrum and because it does not pump protons. Concerning the open question on the proton pump mechanism, it is therefore not too important that all existing X-ray models of L<sup>16,19–22</sup> are at variance with each other concerning the structure of the chromophore and concerning the locations of the internal water molecules. Thorough discussions of the experimental and refinement problems, which cause these differences and similarly apply also to X-ray models of other cryogenically trapped BR intermediates, are given in refs 15 and 16. Note that theoretical treatments, which selected one of these trapped “L” structures as their starting point for simulations, identified some of the X-ray models for the chromophore as incompatible with quantum chemistry.<sup>18,28</sup> Note furthermore that other theoretical treatments<sup>29</sup> took the low-temperature X-ray structures of the various intermediates for granted and derived a mechanistic explanation of the proton pumping from these structures.

We conclude that only time-resolved techniques addressing the BR photocycle at physiological temperatures can provide a valid access to the key mechanistic question of the chromophore

structure(s) in L. These techniques must be capable to cope with the likely conformational heterogeneities discussed in part I of this work<sup>1</sup> (DOI 10.1021/jp902428x) which suggest (at least) two different L states.

Up to now, the only detailed information on the chromophore structure in the L intermediate of the BR photocycle comes from time-resolved vibrational spectroscopy.<sup>10,14,25,30</sup> Therefore, currently, the only access to the chromophore structure in L (and, thus, to an insight into the physical mechanism of proton pumping) can be gained through a high-quality theoretical analysis of vibrational spectra. This view of the issue has inspired this and previous<sup>31</sup> work.

#### Computing the Vibrational Spectra of the Chromophore.

As explained in the Introduction to part I (DOI 10.1021/jp902428x), a vibrational analysis, which aims at a reliable decoding of the observed FTIR and resonance Raman (RR) spectra in terms of the underlying chromophore structures, is a particularly hard task as long as the chromophore is protonated<sup>32</sup> like it is, e.g., in BR<sub>568</sub> and in the early K and L intermediates. Therefore, early attempts<sup>33–35</sup> to model the intramolecular force field of an RSBH<sup>+</sup> empirically with fitted force constants, whose initial values were derived for unprotonated and strongly alternating polyenes, are expected to be error-prone. Nevertheless, all early (and nowadays still accepted) conclusions on the isomeric state of the BR chromophore in the various photocycle intermediates were derived from such analyses (see, e.g., ref 6). Other early descriptions, which were obtained by combining semiempirical quantum chemistry for the RSBH<sup>+</sup> with simplified one-counterion models for the protein environment,<sup>36</sup> suffered from the insufficient accuracy of the employed Hartree–Fock approximation and from the crudeness of the one-counterion model. These difficulties led to unresolved conflicts<sup>9–12</sup> in the structural interpretation of the spectra.

Today, however, DFT/MM techniques are available, which enable much more accurate descriptions of chromophore vibrational spectra (for a review, see ref 37), if the associated protein structures are precisely known. For light-adapted BR<sub>568</sub>, the X-ray structure 1C3W has been gained by Luecke et al.<sup>38</sup> from crystals scattering to a resolution of 1.55 Å. This cryogenic structure is unaffected by the problems of partial occupancy shedding doubts on the quality of intermediate structures.<sup>15</sup> As shown in part I of this work (DOI 10.1021/jp902428x), this reliable starting structure 1C3W has enabled the generation of a room-temperature ensemble of BR structures by MD simulation. Furthermore, one may speculate whether the progress of computer technology and of computational methods can enable a simulation-based construction of accurate models for the K and L intermediates in the near future. Then, one could compute the associated vibrational spectra and compare with observations to conclude on structures. The reliability of such conclusions critically depends on the quality of the employed computational methods, of course.

This is the motivation why we wanted to check whether current DFT/MM hybrid methods can describe the vibrational spectra of the BR<sub>568</sub> chromophore *in situ* with an accuracy that could suffice to decide such intricate questions like that of the chromophore structure in the L intermediate. Furthermore, the DFT/MM results will allow us to judge the quality of earlier attempts for a vibrational analysis of the BR<sub>568</sub> chromophore, which were either based on empirical<sup>33–35</sup> or on simple quantum chemical<sup>36</sup> methods. Because the generally accepted concepts on the isomeric state of the chromophore in the various intermediate states of BR (or of other retinal proteins) originally

derive from the quoted empirical analyses, a corresponding reevaluation seems appropriate.

## Methods

IR line spectra were calculated for the BR chromophore by conventional DFT/MM normal-mode analyses using the all-atom model 1C3W/MD<sup>1</sup> derived from the 100 K X-ray structure 1C3W.<sup>38</sup> In these calculations, three different partitions of the simulation system into DFT and MM fragments were used. Figure 2 depicts and labels these partitions. In the partitions S and M, the two fragments are connected by only one covalent C–C linkage which is treated by the “scaled position link atom method (SPLAM)”.<sup>31</sup> In L, two SPLAM cuts are applied to the C<sub>α,215</sub>–C<sub>215</sub> and C<sub>α,216</sub>–C<sub>216</sub> bonds within the backbone. An additional DFT normal-mode analysis was applied to the isolated DFT fragment of partition M.

All DFT calculations were carried out with the plane-wave code CPMD<sup>39</sup> using the gradient-corrected exchange functional of Becke,<sup>40</sup> the correlation functional of Perdew,<sup>41</sup> and the norm-conserving pseudopotentials of Troullier and Martins.<sup>42</sup> The rectangular box containing the grid for the plane-wave expansion of the Kohn–Sham orbitals was placed around the respective DFT fragment in such a way that no atom of the fragment came closer than 3 Å to one of the faces and that the box volume became minimal. The size of the basis set was then defined by a plane-wave cutoff of 70 Ry. This particular DFT approach will be denoted as “MT/BP”.

Inhomogeneously broadened IR spectra were calculated in the DFT/MM setting by applying the protocol for instantaneous normal-mode analyses (INMA) described in ref 43 to many snapshots taken from the MD trajectories discussed in part I of this work (DOI 10.1021/jp902428x).<sup>1</sup> These DFT/MM computations were carried out with the program combination EGO/CPMD originally introduced in ref 31. If not stated otherwise, the MM fragment was described with the PBR force field.<sup>1</sup>

For comparisons with experimental fundamentals, all harmonic frequencies calculated by MT/BP in the DFT and DFT/MM settings were scaled by a factor of 1.005. This factor is somewhat smaller than the factor of 1.0122 suggested in ref 44 for comparing harmonic MT/BP descriptions of quinone spectra with observed fundamentals.

We will characterize the geometry of the polyene chain within the RSBH<sup>+</sup> by three local observables, which we call the alternation  $\alpha_i$ , the curvature  $\kappa_i$ , and the helicity  $\eta_i$ . Here,  $i$  is the number of the carbon atom  $C_i$  within the RSBH<sup>+</sup> (cf. Figure 1). The local alternation

$$\alpha_i = (-1)^{i-1}(\lambda_i - \bar{\lambda}), \quad i \in \{6, \dots, 15\} \quad (1)$$

measures how much the length  $\lambda_i$  of the C<sub>*i*-1</sub>–C<sub>*i*</sub> bond deviates from the average C–C bond length

$$\bar{\lambda} = \frac{1}{10} \sum_{i=6}^{15} \lambda_i \quad (2)$$

For a normal polyene or an unprotonated retinal Schiff base (RSB),  $\alpha_i$  is positive and assumes everywhere similar values, whereas, for a RSBH<sup>+</sup>, it should decrease from the cyclohexene ring toward the Schiff base.<sup>45</sup> The local curvature

$$\kappa_i = (-1)^i(\varphi_i - 120^\circ) \quad (3)$$

gives the deviation of the bond angle  $\varphi_i$  between the atoms C<sub>*i*-1</sub>–C<sub>*i*</sub>–X<sub>*i*+1</sub> from the value of 120° (X ∈ {C, N}). For regularly extended polyenes,  $\kappa_i$  vanishes everywhere, whereas a total positive curvature

$$\kappa_{\text{tot}} = \sum_{i=6}^{15} \kappa_i \quad (4)$$

is associated with the well-known banana shape of retinal chromophores. Finally, the local helicity

$$\eta_i = (-1)^i(\theta_i - 180^\circ) \quad (5)$$

measures the deviation of the dihedral angle  $\theta_i$  at the C<sub>*i*</sub>–X<sub>*i*+1</sub> bond from the value of 180°, signifying planarity. The polyene helicity

$$\eta_{\text{pol}} = \sum_{i=7}^{15} \eta_i \quad (6)$$

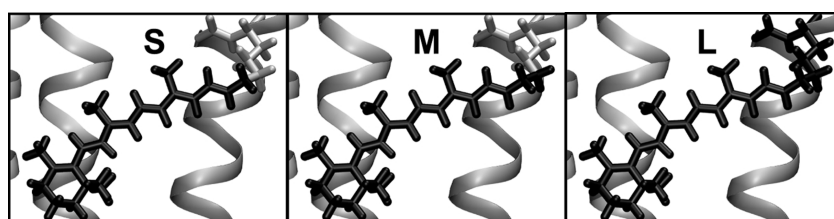
then provides a rough measure of how much the plane of the C=N group is twisted with respect to that of the C<sub>7</sub>=C<sub>8</sub> group attached to the cyclohexene ring.

In the BR chromophore, a positive unit charge is delocalized over the molecule. A clear representation of this delocalization is obtained, if the atomic ESP charges<sup>31,46</sup> resulting from DFT/MM calculations for the respective DFT fragment are summed up over the atom groups  $j$  defined by Figure 3. The total charge in the Schiff base region of the molecule is then given by the quantity

$$q_{\text{SB}} = q_{13-14} + q_{15-N} + q_{\text{Lys}} \quad (7)$$

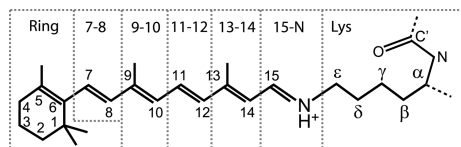
## Results and Discussion

Before comparing experimentally observed spectra with the inhomogeneously broadened IR spectra predicted for the BR<sub>568</sub> chromophore at room temperature by the INMA approach in

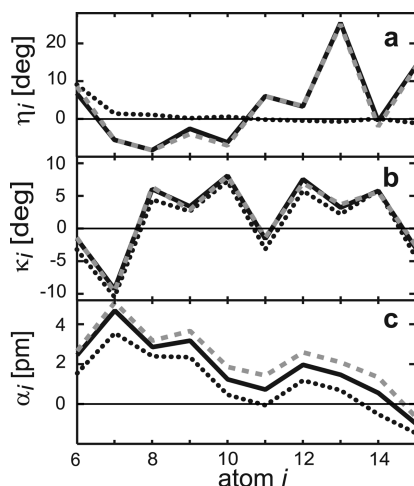


**Figure 2.** Three partitions S, M, and L of the BR simulation model defining differently extended DFT (black) and MM fragments (light gray).





**Figure 3.** The total charges  $q_j$  within the seven indicated atom groups  $j \in \{\text{Ring}, 7-8, 9-10, 11-12, 13-14, 15-\text{N}, \text{Lys}\}$  are used to characterize the charge delocalization predicted by DFT or DFT/MM treatments for the chromophore;  $q_j$  are the sums of the ESP charges<sup>31,46</sup> at the DFT atoms within the respective groups.



**Figure 4.** DFT and DFT/MM results explaining the effects of the apoprotein (structure 1C3W/MD, force fields PBR and C22, DFT fragment M) on the chromophore's geometry. Shown are the local (a) helicity  $\eta_i$  (eq 5), (b) curvature  $\kappa_i$  (eq 3), and (c) alternation  $\alpha_i$  (eq 1) along the polyene chain. Different apoprotein models are distinguished by line style: VAC, black dotted; PBR, black solid; C22, gray dashed.

the DFT/MM setting, we will first show how the descriptions change with increasing complexity and realism of the modeling. Instead of the large structural ensembles resulting from MD simulations at 300 K required for INMA, we will use a single BR structure for these purposes. Thus, we will apply conventional DFT/MM normal-mode analyses to our all-atom model 1C3W/MD<sup>1</sup> of the 100 K X-ray structure 1C3W.<sup>38</sup>

**How the Protein Modifies the RSBH<sup>+</sup>.** To obtain a first insight into the effects of the apoprotein on the properties of the chromophore, we carried out three different normal-mode analyses. The first calculation (labeled as VAC) treated the isolated DFT fragment of the DFT/MM partition M (cf. Figure 2), i.e., the all-*trans* ethyl-RSBH<sup>+</sup> in the vacuum. The two other calculations were executed in the DFT/MM setting for partition M and, correspondingly, included the apoprotein. As a molecular model, we took in both cases the all-atom structure 1C3W/MD, but in one case (labeled as C22), we employed the CHARMM22 partial charges<sup>47</sup> for the apoprotein and in the other case (labeled as PBR) those of the polarized force field PBR.

**Chromophore Geometry.** Figure 4 compares the helicity, curvature, and bond alternation obtained from the VAC, PBR, and C22 calculations. According to the dotted line in Figure 4a, the helicity of the isolated RSBH<sup>+</sup> vanishes nearly everywhere, indicating that this molecule has an almost planar  $\pi$  electron system. The single nonzero value  $\eta_6$  marks a 10° torsion around the C<sub>6</sub>—C<sub>7</sub> single bond connecting the cyclohexene ring with the planar ( $\eta_{\text{pol}} = 0^\circ$ ) polyene moiety of the chromophore. This twist is caused by sterical interactions of the two methyl groups at C<sub>1</sub> with the hydrogen atom at C<sub>8</sub> (cf. Figure 1). As witnessed by the almost identical C22 and PBR helicities, the

embedding of the RSBH<sup>+</sup> into the protein causes twists around nearly all bonds within the polyene moiety which are essentially independent of the detailed electrostatics model. The largest twists are assigned to the C<sub>13</sub>=C<sub>14</sub> and C<sub>15</sub>=N double bonds, resulting in a positive overall helicity of the chromophore's polyene chain ( $\eta_{\text{pol}} \approx 24^\circ$ ). Solely at the C<sub>14</sub>—C<sub>15</sub> single bond, the carbon chain is predicted as locally planar. This result seems to contradict the solid-state NMR data of Lansing et al.,<sup>48</sup> who determined for the H—C<sub>14</sub>—C<sub>15</sub>—H dihedral angle the small but nonzero value of  $16 \pm 4^\circ$ . However, additionally extracting this dihedral angle from the C22 and PBR calculations yields the nonzero values of 13 and 11°, respectively. The smallness of the twist around the C<sub>14</sub>—C<sub>15</sub> bond fits well with the small intensity of the 15-Hoop band at 976 cm<sup>-1</sup> in the IR spectrum of BR<sub>568</sub> in D<sub>2</sub>O.<sup>12</sup>

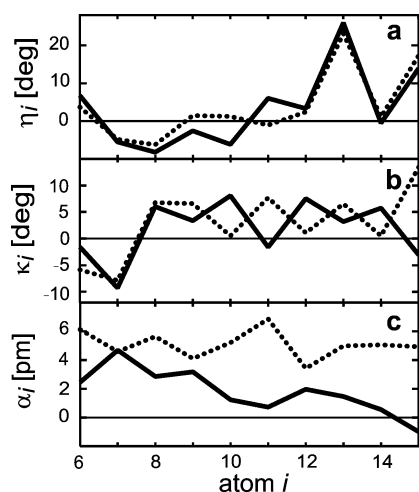
Thus, the various twists around the C—C bonds are mainly caused by steric interactions of the RSBH<sup>+</sup> with its binding pocket. In contrast, the net curvature of the polyene chain predicted by the curves in Figure 4b is mainly caused by intramolecular steric interactions of the two bulky methyl groups at C<sub>9</sub> and C<sub>13</sub> with the neighboring hydrogen atoms. Correspondingly, the local curvatures are particularly large at  $i \in \{8, 9, 10\}$  and  $i \in \{12, 13, 14\}$ . Furthermore, they nicely display the local symmetry of the chromophore structure defined by a mirror plane vertically crossing the polyene chain at C<sub>11</sub> (cf. Figures 4b and 1). Within the protein, the total curvatures  $\kappa_{\text{tot}}$  measure 18.4° for C22 and 18.6° for PBR. Thus, they are independent of the electrostatics model. These total curvatures are much larger than the 6.9° found for the isolated chromophore, showing that steric protein—chromophore interactions additionally contribute to  $\kappa_{\text{tot}}$ .

The bond alternation defined by eq 1 is a very important quantity because it correlates with the stiffness of the various C—C bonds along a polyene chain with respect to torsion and stretching. Here, large  $\alpha_i$  indicate stiff double and floppy single bonds, whereas vanishing  $\alpha_i$  point to bonds of intermediate stiffness. Because the IR spectra of polyene derivatives are dominated by the C—C stretching vibrations, the inspection of a graph like Figure 4c can give a first hint on the relative sizes of the local stretching force constants entering corresponding normal modes.<sup>36</sup>

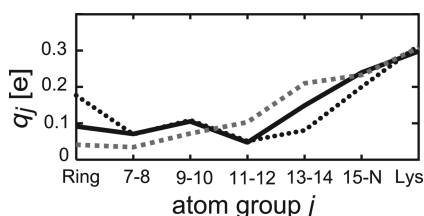
Figure 4c shows the decrease of the bond alternation  $\alpha_i$  from the cyclohexene ring toward the Schiff base which is to be expected for protonated chromophores.<sup>45</sup> The average alternation is smallest for the isolated dye, becomes larger upon transfer into the protein (black solid and gray dashed lines), and sensitively depends on the details of the electrostatics model because for C22 the  $\alpha_i$  are sizably larger than for PBR. In all cases, the local alternation  $\alpha_i$  becomes slightly negative at C<sub>14</sub>, indicating that all three chromophore models tend to assign a somewhat larger double bond character to the C<sub>14</sub>—C<sub>15</sub> bond than to the formal C<sub>13</sub>=C<sub>14</sub> double bond. This enhanced double bond character can partially explain why the twists induced by sterical strains into the chromophore are larger at the C<sub>13</sub>=C<sub>14</sub> than at the C<sub>14</sub>—C<sub>15</sub> bond (cf. Figure 4a).

The DFT/MM calculations start at the chromophore geometry given by the X-ray structure 1C3W.<sup>38</sup> Subsequently, they minimize the total energy function of the hybrid model for BR by slightly varying all atomic coordinates. Because of the limited resolution of the X-ray data and of the high quality of DFT/MM structure calculations, the thus obtained chromophore geometry should be much more accurate than the X-ray model.

Figure 5 illustrates the geometry changes resulting from a DFT/MM energy minimization. The largest changes are ob-



**Figure 5.** X-ray structure 1C3W<sup>38</sup> of the chromophore (dotted lines) compared with the DFT/MM prediction (solid lines, copied from Figure 4) on the chromophore's geometry: (a) helicity; (b) curvature; (c) alternation.

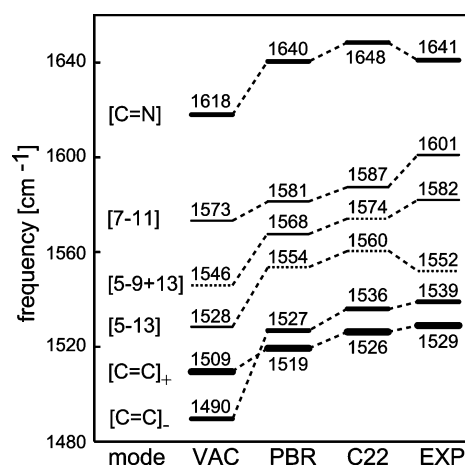


**Figure 6.** Effects of the apoprotein on the delocalization of the positive charge within the chromophore as measured by the group charges  $q_j$  defined in Figure 3. See the caption to Figure 4 for further information.

served for the bond alternation in Figure 5c. Whereas the X-ray model (dotted curve) assigns uniformly a large and positive alternation to the polyene chain, the DFT/MM result (solid curve) yields the expected decrease of alternation toward the Schiff base. Concerning the alternation, the X-ray model reflects the properties of the standard MM force field employed during refinement which imposes short lengths (1.35 Å) to formal double and large lengths (1.46 Å) to formal single bonds. The X-ray curvature in Figure 5b neither shows the banana shape nor the local mirror symmetry at C<sub>11</sub> imposed by the methyl groups at C<sub>9</sub> and C<sub>13</sub>. The corresponding errors of the X-ray model are about 3% and are of the same order of magnitude as the errors of the bond lengths. Solely for the helicity (Figure 5a), the differences are smaller. We suggest that the noted shortcomings of the X-ray structure reflect the prejudices contained in the MM force field used for refinement.

**Charge Distribution within the Chromophore.** Up to this point of the presentation, the bond alternation was the only chromophore property for which we have identified a strong dependence on external electric fields. A second and intimately related property of this kind is the distribution of the positive charge within the chromophore displayed by Figure 6.

Like the local alternations  $\alpha_i$ , also the  $q_j$  show strong variations not only upon transfer of the RSBH<sup>+</sup> into the protein but also upon small changes of the local electrostatics. According to Figure 6, the delocalization of the charge toward the ring is strongest in the vacuum case (black dotted curve) and is successively reduced for the PBR (black solid curve) and C22 (gray dashed curve) models of the electrostatics. Conversely, the charge  $q_{SB}$  remaining in the Schiff base region (cf. eq 7) is smallest for VAC (0.6  $e$ ), intermediate for PBR (0.7  $e$ ), and

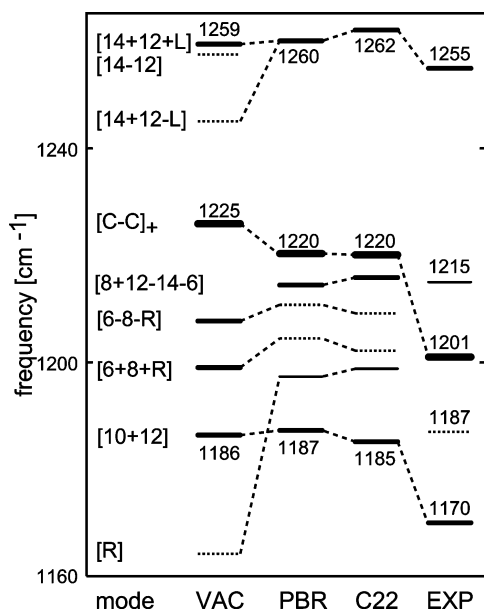


**Figure 7.** Effect of the apoprotein on the C=C and C=N stretch frequencies of the RSBH<sup>+</sup>. The results of the DFT calculation VAC on the isolated RSBH<sup>+</sup> are compared with the DFT/MM calculations C22 and PBR on the BR chromophore and with experimental data (EXP) obtained for BR<sub>568</sub> by Alshuth<sup>49</sup> and by Gerwert.<sup>8</sup> For further explanations, see the text and the caption to Figure 4.

largest for C22 (0.8  $e$ ). A comparison with Figure 4c now shows that  $q_{SB}$  is actually correlated with the size of the bond alternation.

**Vibrational Spectra of the Chromophore.** Because the bond alternation  $\alpha_i$  represents a measure for the C—C stretching force constants in retinal chromophores, the prominent C—C stretching bands in the vibrational spectra of these dyes are expected to show a similarly strong field dependence. Figure 7 clearly verifies this expectation for the high-frequency spectral range characteristic for the C=C and C=N double bond stretches. As compared to the experimental band positions determined by RR spectroscopy for BR<sub>568</sub> at physiological temperatures,<sup>49</sup> the C=C and C=N stretch frequencies calculated for the isolated chromophore show an average redshift of 31.6 cm<sup>-1</sup>. The inclusion of the protein electrostatics reduces this root-mean-square deviation to 11.8 cm<sup>-1</sup> for PBR and to 8.1 cm<sup>-1</sup> for C22, respectively. The quoted deviations depend, of course, on the assignment of observed RR band positions to normal-mode frequencies which is given in Figure 7 by dashed lines. The foundation of this assignment will be discussed further below using data on isotopically labeled chromophores. Note that the increase of the double bond stretching frequencies displayed by Figure 7 for the sequence VAC, PBR, and C22 nicely correlates with the correspondingly increasing bond alternation and localization of positive charge near the Schiff base discussed further above.

Besides frequencies, Figure 7 contains two types of additional information: (I) The thickness of the lines marking the vibrational levels coarsely measures the associated IR intensity. The experimental intensity information required for the drawing of column EXP has been extracted from a FTIR difference spectrum (BR-M) published by Gerwert.<sup>8</sup> (II) The symbols [...] in the first column schematically characterize the compositions of the normal modes resulting from the DFT and DFT/MM treatments. For instance, the symbol [5-13] characterizes a normal mode predominantly composed of an out-of-phase combination of the C<sub>5</sub>=C<sub>6</sub> and C<sub>13</sub>=C<sub>14</sub> stretches. Similarly, the symbol [C=C]<sub>+</sub> denotes a strongly delocalized, in-phase stretching motion of all C=C double bonds in the BR chromophore.<sup>56</sup> Figure 17 in the Supporting Information (SI) explains the given mode symbols through a graphical representation of the various normal modes. According to our experience, the



**Figure 8.** Effect of the apoprotein on the frequencies of C—C single bond stretches of the RSBH<sup>+</sup> (cf. the caption to Figure 7).

mode compositions of the double bond stretches are very stable properties and hardly change for different environments.

With these explanations, the experimental data and the DFT/MM results collected in Figure 7 clearly illustrate the well-known fact that the double bond stretching region in BR's IR spectrum is dominated by three intense bands. The most intense band at 1529 cm<sup>-1</sup> belongs to the totally symmetric ethylenic stretch [C=C]<sub>+</sub>. This band has a large shoulder which is centered at 1539 cm<sup>-1</sup> and belongs to the totally antisymmetric mode [C=C]<sub>-</sub>. According to the DFT/MM calculations, the intensity of the [C=C]<sub>-</sub> mode is by about a factor of 2 smaller than that of the [C=C]<sub>+</sub> mode. Quite intense is also the band of the C=N stretch at 1641 cm<sup>-1</sup>. Smaller IR intensities are predicted for the normal modes [7-11] and [5-9+13] assigned to RR bands at 1601 and 1582 cm<sup>-1</sup>, and a very small IR intensity for the mode [5-13] assigned to a RR band at 1552 cm<sup>-1</sup>.<sup>49</sup>

According to the computational results in Figure 7, the transfer of the RSBH<sup>+</sup> from the vacuum into a protein environment does not only cause an overall blueshift of the frequencies but also has one additional consequence: Whereas in the isolated cation the less intense [C=C]<sub>-</sub> mode is predicted as the lowest C=C stretch, the DFT/MM calculations consistently claim that in BR<sub>568</sub> the more intense symmetric combination [C=C]<sub>+</sub> should be found at a lower frequency. This claim agrees with the available experimental evidence.<sup>35</sup>

As noted above, the frequencies of the double bond stretches exhibit a clear correlation with the field-dependent bond alternation. Thus, one would naively expect an anticorrelation for the frequencies of the single bond stretches. Figure 8 compares the results of the calculations VAC, PBR, and C22 with experimental data<sup>8,49</sup> for the spectral region between 1160 and 1260 cm<sup>-1</sup> which is associated with C—C single bond stretches and harbors the so-called fingerprint bands of the BR<sub>568</sub> chromophore. In the IR spectrum, the fingerprint of the BR<sub>568</sub> chromophore consists of a very intense band at 1201 cm<sup>-1</sup> and of two intense bands at 1170 and 1255 cm<sup>-1</sup>, respectively. In general agreement with the FTIR data, also the calculations predict one very intense and two intense C—C single bond stretches which are thus assigned to the observed fingerprint

bands. Figure 18 in the SI graphically explains the mode compositions of these intense C—C stretches.

Among the C—C stretches generating the fingerprint motif, only the frequency of the most intense mode [C—C]<sub>+</sub> (which is a delocalized in-phase combination of all C—C stretches within the polyene moiety of the RSBH<sup>+</sup>) exhibits the expected anticorrelation with the bond alternation. Correspondingly, the [C—C]<sub>+</sub> frequency is redshifted with increasing alternation by 5.6 and 0.2 cm<sup>-1</sup> in the transitions VAC to PBR and PBR to C22, respectively. Although bringing the computational results closer to the observations, the DFT/MM calculations still overestimate the C—C (Figure 8) and underestimate the C=C frequencies (Figure 7), indicating that the DFT/MM models still underestimate the bond alternation of the chromophore in BR<sub>568</sub>. In the chromophore of BR<sub>568</sub>, the positive charge seems to be even more strongly localized toward the Schiff base than in the DFT/MM models discussed so far.

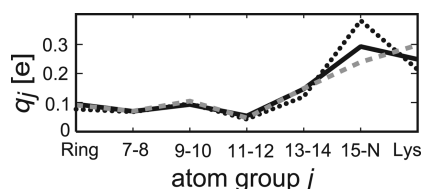
The frequencies of the other C—C stretches do not show a likewise simple dependence on the bond alternation. In the VAC to PBR transition, for instance, the frequency of the mode [14+12+L] increases by 1 cm<sup>-1</sup>, although the C<sub>14</sub>—C<sub>15</sub> and C<sub>12</sub>—C<sub>13</sub> bonds become elongated in this transition, indicating smaller stretching force constants. Such a more complicated behavior has to be expected for a congested spectral region like the one covered by Figure 8, because here the C—C stretches are strongly coupled to many higher and lower frequency modes (including the in-plane bending modes of the hydrogen atoms attached to the polyene chain). On the one hand, this coupling can strongly diminish the effects of altered C—C stretching force constants. On the other hand, in the case of nearly degenerate levels, even small changes of the geometry or of external fields can easily induce large changes of mode compositions and of spectral locations. For instance, in the transition from VAC to PBR, one sees a large blueshift of a C—C stretching mode localized in the cyclohexene ring (mode [R]) and the appearance of a new mode [8+12-14-6] spectrally near the [C—C]<sub>+</sub> mode which was absent in VAC. The complexity of the coupling patterns in the densely populated spectral region of the C—C single bond stretches is the reason why we will mainly look at the much simpler high-frequency region of the C=C and C=N double bond stretches in our continued assessment of different theoretical descriptions.

**Intermediate Summary.** The comparison of the single point calculations VAC, PBR, and C22 with experimental data has shown that the inclusion of the apoprotein not only enables a highly precise determination of the chromophore geometry but also generally brings the calculated frequencies much closer to the experimental values. On the other hand, the DFT/MM calculations seem to overestimate the delocalization of the positive charge into the polyene moiety. In this respect, the C22 force field apparently performs better than PBR which is an unexpected finding. However, the PBR force field has been developed for a different partitioning (L) of the system into DFT and MM fragments. Thus, we now must check how the choice of the DFT fragment affects the descriptions.

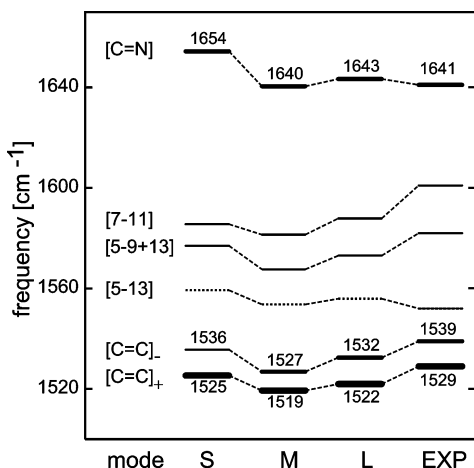
**Effects of the DFT Fragment Size.** For this study, we have chosen the three partitionings S, M, and L of the system into DFT and MM fragments (cf. Figure 2), because we wanted to keep the DFT fragment for reasons of computational feasibility as small as possible. Here, the apoprotein was modeled by PBR<sup>1</sup> and the water molecules by Jorgensen's transferable three-point potential (TIP3P).<sup>50</sup>

As demonstrated by Figure 19 in the SI, the variation of the DFT fragment size has a negligible effect on the curvature of





**Figure 9.** Effect of the DFT fragment size (cf. Figure 2) on the charge distribution  $q_j$  within the  $\text{RSBH}^+$  (DFT/MM results, structure 1C3W/MD force field PBR): S, black dotted; M, gray dashed; L, black solid.

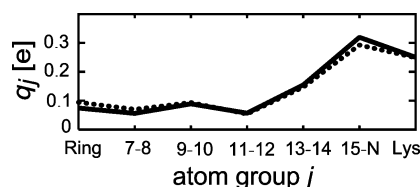


**Figure 10.** Effect of fragment size S, M, or L on the C=C and C=N stretch frequencies of the BR chromophore from DFT/PBR computations on the structure 1C3W/MD. Like in Figure 7, the computed line spectra are compared with experimental data.

the polyene moiety and causes only small changes of the dihedral angles and of the bond alternation near the Schiff base. In good correlation with the charge distribution shown in Figure 9, the bond alternation is only slightly larger for partition S than for M or L. In fact, for S, the total charge  $q_{\text{SB}}$  near the Schiff base (cf. eq 7) is only by 3% larger than that for M or L. Figure 9 confirms that the variation of the DFT fragment changes the charge distribution solely near the C=N group.

The slightly larger value of  $q_{\text{SB}}$  leads for model S to the expectation of somewhat stiffer C=C and C=N double bonds. Figure 10 demonstrates that this is actually the case. Compared to M and L, the double bond stretching frequencies obtained for S are shifted to the blue by averages of 8 and 4  $\text{cm}^{-1}$ , respectively. However, as compared to model M, also for L the frequencies are shifted toward the blue. As a result, the root-mean-square deviation between the experimental and calculated frequencies is reduced from 11.8  $\text{cm}^{-1}$  for M to 9.3  $\text{cm}^{-1}$  for S and to 7.8  $\text{cm}^{-1}$  for L. The latter deviation is even a little smaller than the value of 8.1  $\text{cm}^{-1}$  noted above for combining the DFT fragment M with the C22 force field. Correspondingly, it seems that the PBR force field in combination with the largest DFT fragment L provides the best description of the experimental data.

Note, however, that the model S with its larger alternation and charge  $q_{\text{SB}}$  performs better than L when one exclusively considers the C=C stretch frequencies. It is only the 13  $\text{cm}^{-1}$  overestimate of the C=N stretch frequency found for S which excludes this most simple modeling. Replacing the lysine chain by a methyl group, as one does in the DFT treatment with fragment S, apparently places the boundary between the DFT and MM regions too close to the Schiff base group for a correct evaluation of the C=N force constant. Furthermore, the suboptimal performance of model M is caused by the choice



**Figure 11.** Distribution of charges  $q_j$  within the chromophore for conformation  $C_1$  and DFT/MM partition L: The 300 K ensemble average (solid curve) is compared with the 100 K result (dotted curve). See the text for an explanation.

of the electrostatics for the lysine chain in the PBR force field favoring a somewhat too strong delocalization of the positive charge. As a result, the computationally most expensive variant L among the studied alternatives for DFT/MM partitioning had to be chosen for the subsequent study of temperature effects.

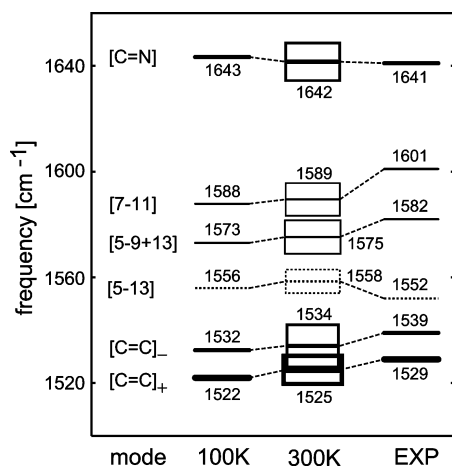
**Temperature Effects.** The above descriptions of chromophore structures and IR spectra were obtained for an all-atom version of the X-ray structure 1C3W.<sup>38</sup> Correspondingly, these computational results pertain to a system at 100 K, whereas the experimental spectra employed for comparisons were recorded at room temperature.

In part I of this work (DOI 10.1021/jp902428x), we succeeded to generate a 300 K conformational ensemble of the chromophore binding pocket by MD simulations.<sup>1</sup> Due to the use of the polarized force field PBR, these simulations preserved the crystallographic structure over extended time spans (>50 ns). Using a specially designed sampling technique, the 300 K ensemble was shown to be composed of (at least) two, and possibly even three, distinct and slowly interconvertible conformational substates called  $C_1$ ,  $C_2$ , and  $C_3$ , respectively. The increase of the temperature from 100 to 300 K turned out to have two effects: (i) Within a given substate, the amplitudes of the thermal fluctuations become larger. (ii) The relative populations of the substates  $C_i$  are modified. Whereas the crystallographic conformation  $C_1$  is predominantly populated at 100 K, conformation  $C_2$  gains a considerable statistical weight at higher temperatures. For the rare conformation  $C_3$ , the limited statistics sampling precluded an accurate prediction of its weight at 300 K. We expect that the effects (i) and (ii) will modify the average structure and the IR spectra of the chromophore.

**Effects of Thermal Fluctuations.** First, we will address point (i) for conformation  $C_1$ . Our all-atom model 1C3W/MD obviously represents a 100 K snapshot of  $C_1$ . Furthermore, a PBR-MD simulation starting at 1C3W/MD has rendered for  $C_1$  a 300 K trajectory covering 50 ns. Applying the INMA approach (cf. Methods) to this trajectory enables the computation of room-temperature IR spectra. Thus, for state  $C_1$ , one can identify the effects of increasing thermal fluctuations by comparing average 300 K properties of the chromophore with data from the 100 K snapshot. At 100 K, the thermal fluctuations are small and, therefore, a single minimized structure suffices.

Figure 11 compares for partition L and conformation  $C_1$  the average distribution of charges  $q_j$  within the chromophore at 300 K with the distribution within the 100 K snapshot (dotted curve, identical to the solid curve in Figure 9). The 300 K average has been calculated from 25 snapshots selected from the central part of the 50 ns PBR-MD trajectory at 5 ps temporal increments. This snapshot ensemble has also been used for the construction of all other DFT/MM results that will be presented for  $C_1$  at 300 K.

Figure 11 shows for state  $C_1$  that elevating the temperature from 100 to 300 K leads within the chromophore to a slight shift of positive charge toward the Schiff base. Correspondingly, the value of  $q_{\text{SB}}$  increases by 3% upon heating and becomes as



**Figure 12.** Average C=C and C=N stretch frequencies and corresponding band widths in conformation  $C_1$  at 300 K compared with the 100 K results (structure 1C3W/MD, partition L; see also Figure 10). The band widths are indicated by heights of boxes measuring the standard deviations of the underlying 300 K frequency ensembles. Experimental data (cf. Figure 7) are given as a reference.

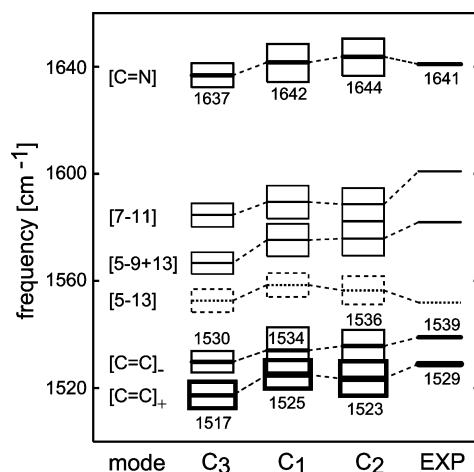
large as in the low-temperature fragment S. In line with the behavior of  $q_{SB}$ , also the average bond alternation increases a little (0.1 pm) upon heating. Figure 20 in the SI documents that these and other structural changes induced by heating are actually quite small. The noted increases of  $q_{SB}$  and of the bond alternation entail for state  $C_1$  the prediction that heating should shift the C=C stretch frequencies to the blue.

Figure 12 confirms this prediction. Upon heating, all C=C frequencies are shifted nearly uniformly by about 2 cm⁻¹ to the blue. In contrast, the C=N stretch is shifted by 1 cm⁻¹ to the red. Correspondingly, the deviation between the calculated and observed double bond stretching frequencies decreases from 7.8 to 6.6 cm⁻¹. Despite the quite close matching of calculated and experimental frequencies, which meanwhile has been achieved through a successively more realistic and complex modeling of the experimental situation, the C=C stretching force constants still seem to be a little too small. Thus, even our current chromophore model still features a charge delocalization that is apparently a little too large.

**Effects of Conformational Transitions.** To characterize the three conformational substates identified in part I (DOI 10.1021/jp902428x), we have selected, in addition to the 25  $C_1$  snapshots, 20 snapshots from each of the extended 300 K MD trajectories sampling the conformational states  $C_2$  and  $C_3$ .

Figure 21 and the associated text in the SI present, compare, and discuss associated DFT/MM results that explain the influence of the respective conformational state on the structure of the chromophore. Inspecting these data, one concludes that the steric strain exerted by the binding pocket on the chromophore decreases in the sequence  $C_1 \rightarrow C_2 \rightarrow C_3$ . For instance, in state  $C_3$ , the chromophore has a reduced curvature just like the model VAC discussed further above. Furthermore, in  $C_2$  and  $C_3$ , the polyene moiety of the RSBH<sup>+</sup> exhibits much smaller twists than in the sterically restrained but enthalpically favored<sup>1</sup> conformation  $C_1$ . On the other hand, the localization of positive charge near the Schiff base and the bond alternation are seen to increase in the sequence  $C_3 \rightarrow C_1 \rightarrow C_2$  such that the C=C stretching frequencies should exhibit blueshifts in that sequence.

Figure 13 now shows that the C=C stretching frequencies of the states  $C_1$  and  $C_2$  are actually blueshifted with respect to those of state  $C_3$ . However, for the transition from  $C_1$  to  $C_2$ , no



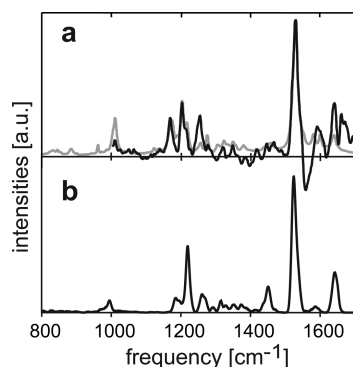
**Figure 13.** Average C=C and C=N stretch frequencies and standard deviations in the 300 K conformations  $C_1$ ,  $C_2$ , and  $C_3$ . See the caption to Figure 12 for further explanations.

overall blueshift is seen, indicating that the spectral positions of the C=C stretches do not exclusively depend on the degree of charge delocalization but also on the chromophore geometry (as mentioned above, in  $C_1$ , the chromophore exhibits larger twists than in  $C_2$ ). Compared with the experimental double bond stretch frequencies, the calculated values show root-mean-square deviations of 11.1 cm⁻¹ for  $C_3$  and of 6.6 cm⁻¹ for  $C_1$  and  $C_2$  (the latter two spectra deviate from each other by 1.6 cm⁻¹).

Considering the band widths, which are measured by the heights of the boxes surrounding the frequency levels, one concludes that the IR bands associated with the three conformational states will strongly overlap. Therefore, particularly the IR spectra of  $C_1$  and  $C_2$  should be very similar in the high-frequency region (and as shown by Figure 22 in the SI, such a similarity also holds in the fingerprint region). The structural differences between  $C_1$  and  $C_2$ , which are mainly localized at the aliphatic chain connecting the RSBH<sup>+</sup> with the polypeptide backbone and at the backbone near Lys216, are thus predicted to be spectroscopically silent.

**Total Spectral Effect of Heating.** According to the results of part I (DOI 10.1021/jp902428x), the conformations  $C_1$  and  $C_2$  should dominate the 300 K conformational ensemble, whereas  $C_3$  should occur in much smaller or even negligible amounts. Thus, we now assume that the crystallographic conformation  $C_1$  is exclusively present at 100 K and that a 42% admixture of  $C_2$  appears at 300 K. Then, the DFT/MM normal-mode analysis of the structure 1C3W/MD can be taken as our description of the IR spectrum of the BR chromophore at cryogenic temperatures (cf. column “L” in Figure 10 or column “100 K” in Figure 12). Furthermore, we can construct an INMA prediction for the room-temperature IR spectrum of the BR chromophore by choosing 25 snapshots from the  $C_1$  trajectory and 18 snapshots from the  $C_2$  trajectory at 300 K to model a 58:42 conformational mixture. Next, one can compare this 300 K spectrum with its 100 K relative to study the combined effects of enhanced thermal fluctuations and changed conformational composition.

The graph thus obtained for the C=C and C=N stretching frequencies would be nearly identical to the one shown in Figure 12, because the frequencies calculated for the 300 K mixture differ from those shown in Figure 12 for the pure conformation  $C_1$  at 300 K by at most 1 cm⁻¹. Because of this similarity, the 300 K mixture data are not documented at this point. Instead, they will be shown and discussed in a different context further below (cf. Figure 15).



**Figure 14.** IR spectrum (b) of the BR chromophore at 300 K as calculated by the INMA procedure in the DFT/MM setting for a 58:42 mixture of conformations  $C_1$  and  $C_2$  compared with experimental spectra. The black curve (a) is a BR-M FTIR difference spectrum extracted from the thesis of Gerwert.<sup>8</sup> The gray curve in the background of part a is a RR spectrum of BR<sub>568</sub> taken from the thesis of Alshuth.<sup>7</sup> All spectra have been normalized to identical heights of the dominating peak of the ethylenic stretch.

Interestingly, however, the 42% admixture of  $C_2$  sizably decreases the differences between the 100 and 300 K data shown in Figure 12 (which refer to pure  $C_1$ ). As a result, in the high-frequency region, the band positions are predicted to shift by at most 2  $\text{cm}^{-1}$  upon heating. Furthermore, the 42% admixture of  $C_2$  reduces the root-mean-square deviation of the calculated from the experimental double bond stretching frequencies slightly from 6.6  $\text{cm}^{-1}$  (pure  $C_1$ ) to 6.5  $\text{cm}^{-1}$  (mixture). These numbers have to be compared with the deviation of 7.8  $\text{cm}^{-1}$  identified for the cryogenic model and suggest that our 300 K mixture model should describe the vibrational spectra of the BR chromophore reasonably well.

**The IR Spectrum of the BR Chromophore at 300 K.** To check this expectation on the quality of our 300 K model, we construct from our INMA data on the conformations  $C_i$  a prediction for the absolute IR spectrum  $I(\nu)$  of the BR chromophore. For this purpose, we center at each INMA frequency a normalized Gaussian with a standard deviation of 5  $\text{cm}^{-1}$  (here, the width has been chosen such that the resulting superposition is sufficiently smooth, despite the limited size of the INMA snapshot ensemble, while still allowing the distinction of neighboring bands). Weighting this Gaussian by the intensity calculated for the associated normal mode, we determine the IR spectrum  $I(\nu|C_i)$  of  $C_i$  by summing over all Gaussian bands which can be derived from the associated INMA data set. For  $C_1$  and  $C_2$ , the INMA data sets contained 25 and 20 line spectra, respectively. The predicted spectrum  $I(\nu)$  is then given as the mixture  $0.58I(\nu|C_1) + 0.42I(\nu|C_2)$ .

Figure 14b shows the resulting prediction for the IR spectrum of BR<sub>568</sub> in the frequency range from 800 to 1700  $\text{cm}^{-1}$ . For visual comparison, Figure 14a also shows a BR-M FTIR difference spectrum and a RR spectrum of BR<sub>568</sub>. These spectra have been scanned from the theses quoted in the caption. Quite obviously, the comparisons of the depicted spectra are hampered by three facts:

(i) The RR and IR intensities follow from different mechanisms. Nevertheless, for BR, the spectra look quite similar with the notable exception of the third fingerprint band at 1255  $\text{cm}^{-1}$  which is intense in IR and weak in RR. (ii) The FTIR spectrum (a) is a difference spectrum between the initial state BR<sub>568</sub> of the photocycle and the long-lived intermediate M. However, because the chromophore is deprotonated in M, the IR intensities of the M chromophore are small. Correspondingly small are the negative bands contributed by the M chromophore to the

difference spectrum (a). (iii) In contrast to RR, FTIR contains additional contributions from the protein. Particularly in the spectral range between 1500 and 1700  $\text{cm}^{-1}$ , the FTIR difference spectrum shows signatures of strong amide bands. These difference bands signify that the protein backbone undergoes conformational changes in the BR to M transition. Thus, RR spectra are necessary to identify the chromophore bands in FTIR difference spectra.

Independent of these difficulties, the main spectral features in the regions of the C—C single and double bond stretching vibrations are clearly visible in the FTIR difference and in the calculated IR spectra: Three fingerprint bands are seen near 1200  $\text{cm}^{-1}$ , and above 1500  $\text{cm}^{-1}$ , one immediately recognizes the strong ethylenic and C=N stretching vibrations. The visual inspection additionally suggests that the calculated band widths are quite close to the observations.

However, one also sees the remaining shortcomings of the description: Whereas FTIR determines for the three fingerprint bands nearly equal intensities, the calculation assigns too small intensities to the two companions of the central and very intense [C—C]<sub>+</sub> band. Furthermore, the calculated fingerprint bands are not as widely spread as the experimental ones and their center is shifted by about 13  $\text{cm}^{-1}$  to the blue.

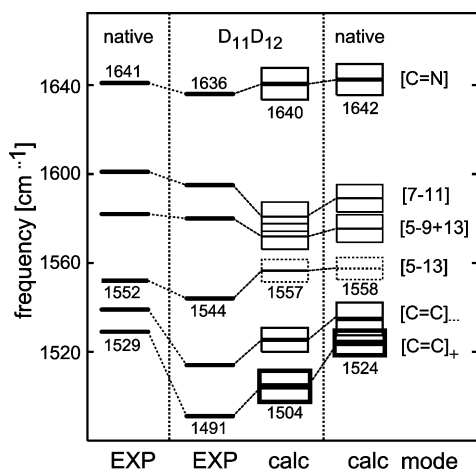
In the discussion of Figure 8, we have suggested that these shortcomings are most likely caused by a delocalization of positive charge which is smaller in the chromophore of BR<sub>568</sub> than in its DFT/MM model. For the model, a reduction of the delocalization requires that the localizing electric field becomes strengthened, e.g., by moving some negative charge density a little closer to the Schiff base group.<sup>36</sup> Therefore, the remaining differences seem to indicate that our electrostatics model is—despite the use of a polarized force field for the apoprotein—not yet good enough.

**The Effect of Polarized Water Molecules.** As explained in part I of this work (DOI 10.1021/jp902428x),<sup>1</sup> we restricted the polarization of the MM fragment to the protein material in the chromophore binding pocket of BR and chose for the three crystallographic water molecules in this pocket the standard nonpolarized TIP3P water model.<sup>50</sup> However, after completion of the MM—MD simulations and of the computation of the associated room-temperature spectra, we found by DFT/MM calculations that the dipole moments of these three water molecules are by about 10% larger than those of water molecules in the bulk liquid.<sup>1</sup>

To estimate the influence of more polar water models on the chromophore spectra, we repeated for the structure 1C3W/MD the DFT/MM calculation “L” of the chromophore’s cryogenic IR spectrum (cf. Figure 10) with such water models. As demonstrated by Figures 23 and 24 in the SI, the larger water dipoles left the C=N stretch frequency invariant and shifted the C=C stretch frequencies on average by 2  $\text{cm}^{-1}$  to the blue and the frequencies of the C—C fingerprint bands by about 1.5  $\text{cm}^{-1}$  to the red. Therefore, the deviations of the calculated stretch frequencies from the experimental data reduced from 7.8 to 7.2  $\text{cm}^{-1}$  for the double bonds and from 15.1 to 13.8  $\text{cm}^{-1}$  for the single bonds (see the SI for further details).

As a result, the use of polarized water molecules should not only further stabilize the native structure in MD simulations<sup>1</sup> but should also further improve the match between the experimental data and the INMA spectra. Transferring, for instance, the 2  $\text{cm}^{-1}$  blue shift of the C=C stretches, which is caused by the larger water dipoles (Figure 23 in the SI), to the room-temperature conformational mixture reduces the frequency deviation from 6.5 to 5.6  $\text{cm}^{-1}$  for the double bonds.





**Figure 15.** Effect of deuteration at  $C_{11}$  and  $C_{12}$  on the  $C=C$  and  $C=N$  stretch frequencies of the BR chromophore. RR frequencies<sup>35</sup> labeled as EXP are compared with DFT/MM results (calc) obtained for our 300 K conformational mixture (0.58  $C_1$  + 0.42  $C_2$ ). The assignments are indicated by the dashed lines and are explained in the text (cf. additionally the caption to Figure 12).

**Accuracy Exemplified for an Isotope Effect.** The remaining deviations between the calculated spectra and the experimental data represent important guidelines for further improvements. Nevertheless, the thus achieved description is not too bad either. Quite certainly, it is much better than the combination of semiempirical quantum chemistry with one-counterion models applied earlier<sup>36</sup> to the vibrational analysis of the BR chromophore or even the purely empirical modeling in ref 35. However, a precise quality assessment requires the consideration of isotope effects. In that respect, the chromophore of BR is a particularly fortunate case because RR spectra on a plethora of isotopically labeled chromophores are available.<sup>35</sup> At the same time, this wealth of data makes the presentation of corresponding results quite cumbersome, because one has to analyze many more details.

Therefore, we decided to focus this paper on the analysis of how the protein binding pocket and the conformational heterogeneity affect the chromophore spectra and to what extent an increasingly more accurate and complex modeling can improve the descriptions. Correspondingly, here we present solely one carefully selected isotope effect. The SI discusses three further isotopic substitutions. One is the spectral effect of deuterating the Schiff base, one the effect of inserting a  $^{13}C_{14}$ ,  $^{13}C_{15}$  double label into the chromophores carbon skeleton, and one is the effect of  $^{15}N$  substitution.

Among all isotopically labeled chromophores investigated by Smith et al.,<sup>35</sup> there is only one species which shows six nonoverlapping and clearly visible RR bands in the high-frequency region of the double bond stretches. This species is the 11,12, dideutero RSBH<sup>+</sup>. Because the chromophore contains six double bonds, the RR spectrum of this compound uniquely defines the frequencies of six double bond stretching modes. All other studied isotopomers<sup>35</sup> show a considerable overlap of bands and, therefore, the determination of all six  $C=C$  frequencies requires error prone deconvolution techniques.

The central columns of Figure 15 contain the RR and DFT/MM results for the double bond stretch frequencies of the 11D,12D-chromophore. Here, the assignment of the calculated to the experimental frequencies was exclusively based on the respective spectral sequences. The modes calculated for the labeled chromophores could be easily assigned to those of native chromophore by visual comparison of mode composi-

tions. The resulting assignment of the labeled to the unlabeled RR frequencies and to the associated modes generally agrees with that by Smith et al.<sup>35</sup>

An exception is the 11D,12D-band at  $1544\text{ cm}^{-1}$  which our DFT/MM calculation assigns to an almost pure  $C_5=C_6$  stretch calculated at  $1557\text{ cm}^{-1}$ . Despite its considerable intensity, Smith et al.<sup>35</sup> had assigned this band, like the much weaker RR band at  $1552\text{ cm}^{-1}$  in the native chromophore, to a combination mode of unknown composition. This judgement was caused by the results of their empirical normal-mode analysis which consistently predicted the  $C_5=C_6$  stretch near  $1600\text{ cm}^{-1}$ . Now, the pattern of isotopic shifts shown in Figure 15 clearly demonstrates that the RR bands at  $1552\text{ cm}^{-1}$  (native) and  $1544\text{ cm}^{-1}$  (11D,12D) both belong to the mode [5-13]. Note that Grossjean et al.<sup>36</sup> arrived at the same conclusion though with much weaker arguments.

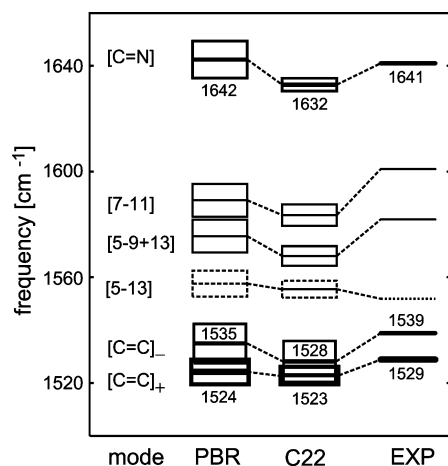
The above discussion is a typical example for the kind of reasoning required in the vibrational analysis of complex spectra. It shows that the assignment of bands to normal modes is greatly simplified if a calculation predicts the various normal modes in the correct spectral sequence and with reasonable compositions. Then, the inevitable differences between the observed frequencies of anharmonic fundamentals and the calculated harmonic frequencies do not necessarily preclude correct assignments.

**Can One Distinguish Conformations?** As outlined in the Introduction, this work was motivated by the question whether current DFT/MM hybrid methods can describe the vibrational spectra of the BR<sub>568</sub> chromophore *in situ* with an accuracy that could suffice to decide such intricate questions like that of the chromophore structure in the L intermediate.

A partial answer to this question has been given by the comparison in Figure 13. The two dominantly populated lysine conformations  $C_1$  and  $C_2$  do not induce significant spectral shifts and, therefore, cannot be distinguished by IR. Here, the lack of effects of the conformational states on the IR spectrum is due to strong similarities of the chromophore structures and of the electrostatics within the binding pocket. In particular, the hydrogen bonded network stabilizing the binding pocket [cf. Figure 2 in part I (DOI 10.1021/jp902428x)] is preserved for both conformations. However, for the purpose of identifying the chromophore structure in L, the DFT/MM descriptions must solely be accurate enough to distinguish strongly different chromophore geometries and hydrogen bonding patterns in the chromophore binding pocket. The latter issue will now be checked.

For this purpose, recall that in part I (DOI 10.1021/jp902428x) we encountered a non-native conformation  $C_{nn}$  of the hydrogen bonded network within the binding pocket.  $C_{nn}$  was determined as the artificial attractor which hampers all MD simulations modeling the apoprotein by a nonpolarizable force field like C22 [cf. part I (DOI 10.1021/jp902428x)].  $C_{nn}$  is characterized by an erroneous hydrogen bond between the RSBH<sup>+</sup> and Asp85. To analyze the spectral effect of this altered conformation, we have selected five snapshots from the C22-MD trajectory documented in Figure 4 of part I (DOI 10.1021/jp902428x) and have calculated an INMA-IR spectrum by DFT/MM. As demonstrated by Figure 27 in the SI, all selected snapshots belong to the non-native conformation  $C_{nn}$ .

Figure 16 shows that all double bond stretching frequencies obtained from the C22-INMA treatment are redshifted with respect to the PBR-INMA frequencies (the latter data were shown already in Figure 15). Due to this redshift, which on average amounts to  $6\text{ cm}^{-1}$ , the root-mean-square deviation between the calculated and experimental double bond stretching



**Figure 16.** Comparison of the INMA double bond frequencies obtained for the PBR force field (43 snapshots: 25  $C_1$ , 18  $C_2$ ) and for the C22 force field (5 snapshots, conformation  $C_{nn}$ ). Experimental data are given as a reference.

frequencies increases from  $6.5\text{ cm}^{-1}$  (PBR) to  $11\text{ cm}^{-1}$  (C22). Here, the  $\text{C}=\text{N}$  double bond stretch shows the largest redshift of about  $10\text{ cm}^{-1}$ .

The just identified redshift is exclusively due to the differences between the conformational states, because a blue shift is expected if one solely replaces the PBR by the C22 force field while leaving the conformational ensemble unchanged. The latter expectation derives from Figures 7 and Figure 26 (in the SI) which both show that the double bond frequencies are shifted on average by  $6\text{ cm}^{-1}$  to the blue, if the description of the apoprotein is switched from PBR to C22 while the underlying structure (1C3W/MD) remains unchanged. The only difference between Figures 7 and 26 (in the SI) is the choice of the DFT/MM partitioning (Figure 7, M; Figures 26 and 16, L).

In summary, the IR spectrum calculated for the  $C_{nn}$  conformation with its non-native hydrogen bonding pattern sizably differs from the spectrum calculated for the native pattern (Figure 28 in the SI compares these spectra within the range  $1100\text{--}1700\text{ cm}^{-1}$ ). Furthermore, the  $C_{nn}$  spectrum deviates much more strongly from the experimental spectrum than the one obtained for our mixture model.

The SI analyzes the causes of these spectral shifts in some more detail. Here, Figure 29 shows the structural changes induced in the chromophore by the transition to the  $C_{nn}$  conformation. Although the charge localization and bond alternation are seen to increase in this transition, these changes do not entail the correspondingly expected blue shift of the double bond stretches, because a modified pattern of twists around  $\text{C}\text{--}\text{C}$  bonds within the Schiff base region of the  $\text{RSBH}^+$  apparently leads to an overcompensating redshift.

These results show that not all conformational changes in the binding pocket of the BR chromophore are spectroscopically silent. Instead, all those changes, which are connected with distinct variations of the local electrostatics and with associated changes of the chromophore geometry, should generate clear signatures in the calculated IR spectra. The results furthermore demonstrate that substantial deviations of the calculated structural ensemble from the experimental one are associated with strongly increased deviations between calculated and experimental frequencies. Consequently, room-temperature ensembles calculated by MD with a conventional nonpolarizable force field, which therefore feature a non-native hydrogen bonded network, are badly suited as starting points for DFT/MM calculations of the chromophore's IR spectra. Note that the latter conclusion

represents an *a posteriori* justification for our tedious approach of first constructing a PBR force field for MD simulations.

The fact that each improvement of the physical modeling (increasing the size of the DFT fragment, accounting for effects of room-temperature fluctuations and of conformational flexibility, including the polarization also of the water molecules, and choosing a proper conformational ensemble through the use of PBR instead of C22) brought the DFT/MM descriptions closer to the experimental findings suggests that, conversely, strong deviations between calculated and experimental spectra can be taken as clues for rating suggested conformations as less likely.

## Summary and Outlook

The above results have made clear that the aim to compute the room-temperature vibrational spectra of the retinal chromophore in  $\text{BR}_{568}$  by DFT/MM hybrid techniques as accurately as possible requires utter technical care.

As a prerequisite, one must generate by MM–MD simulations a quite precise and physically realistic model of the room-temperature conformational ensemble in the chromophore region of the protein. As we have shown in part I of this work (DOI 10.1021/jp902428x),<sup>1</sup> MD simulations that are based on standard nonpolarizable MM force fields like C22 produce a non-native conformation  $C_{nn}$ , which exhibits a strongly altered electrostatics in the surroundings of the  $\text{RSBH}^+$  and is associated with a changed chromophore geometry. Above, we have shown that such deviations from the native conformational ensemble preclude accurate computations of the chromophore's vibrational spectra, thus proving one of our working hypotheses. Because we attributed the apparent instability of the native conformation found in previous MD descriptions to the lacking inclusion of the electronic polarizability and because well-established polarizable force fields for proteins were not yet available, we had to generate the polarized force field PBR by DFT/MM calculations.<sup>1</sup>

The PBR-MD simulations then reproduced the crystallographic conformation  $C_1$  as the most stable state at cryogenic and ambient temperatures. For 300 K, the simulations revealed the existence of a second conformational state  $C_2$  at sizable amounts (42%).<sup>1</sup> As we have shown above, the structural differences of  $C_1$  and  $C_2$  (which are localized in the backbone and in the aliphatic chain linking the  $\text{RSBH}^+$  to the backbone) hardly affect the vibrational spectra of the chromophore.

With a room-temperature conformational ensemble of the chromophore region in  $\text{BR}_{568}$  at hand, the DFT/MM computation of the chromophore's room-temperature IR spectra by applying the INMA procedure had become possible. For this purpose, a suitable partitioning of the system into DFT and MM fragments had to be chosen. Here, the largest DFT fragment L (cf. Figure 2) turned out to be the best choice and the correspondingly enhanced computational effort had to be spent. Similarly, also all further improvements of the modeling, in which initially simplified assumptions were given up in favor of physically more realistic ones, led to an increasingly better description of the well-known chromophore spectra. In these improvements, e.g., the rigid cryogenic X-ray structure was replaced by the 300 K conformational ensemble resulting from our PBR-MD simulations or the standard TIP3P water models were replaced by more polar ones, thus accounting for the corresponding DFT/MM results.<sup>1</sup>

Concerning the structure of the chromophore in  $\text{BR}_{568}$ , the resulting DFT/MM description is much more accurate than the available experimental structures (cf. Figure 5 and the associated

discussion). Concerning the chromophore's vibrational spectra, however, this description still leaves room for further improvements. Here, particularly the restriction to the most simple density functionals, which currently is imposed by the use of the program CPMD<sup>39</sup> as the DFT code in our hybrid program, is a major obstacle. Thus, an extension of our QM/MM interface to a more versatile DFT package (e.g., ref 51) is required for more accurate descriptions.

Nevertheless, the current description is quite clearly good enough to explain the experimental spectra. Evidence for this claim has been given by considering several selected effects of isotope substitutions (in the text and in the SI). Furthermore, such a DFT/MM description is good enough to distinguish non-native from native conformations by comparing calculated spectra with observed ones, if the electrostatics in the chromophore binding pocket and the chromophore geometry are significantly different for these conformations. Also, this second claim has been substantiated by providing an example. Here, it has been shown that replacing the native conformational mixture by the non-native conformation  $C_{nn}$  strongly deteriorates the description of the chromophore's IR spectrum in BR<sub>568</sub>.

These results nourish the hope that DFT/MM analyses of chromophore's vibrational spectra in the early photocycle intermediates K and L can help to identify the associated chromophore structures. As we have suggested in an initial section serving to motivate our work, particularly the analysis of L at room temperature should be capable of shedding light on the yet unknown mechanism of light-driven proton pumping in BR. As long as a highly resolved, room-temperature X-ray structure of L remains elusive, one will have to choose a MD-based modeling approach instead. Here, one will have to start with various assumptions on the primary photoisomerization and will have to carry out subsequent MD simulations extending over microsecond time spans to compute the altered distribution of water molecules characterizing the L state of the protein. Quite clearly, a standard nonpolarizable force field will not be good enough for such MD simulations because these simulations aim at the reliable prediction of the protein conformation in L. One will have to use a polarizable or polarized force field instead. Subsequently, the quality of the conformation computed for the room-temperature L intermediate can be scrutinized by comparing the vibrational chromophore spectra observed at room temperature by time-resolved spectroscopy with DFT/MM predictions. A match of comparable quality as the one achieved in this work for BR<sub>568</sub> would then indicate that the chromophore structure in L has been finally identified. With this structure at hand, the mechanism of proton pumping in BR should be accessible.

The DFT/MM technology used by us in this work is computationally not efficient enough for the purpose sketched above or for standard applications to other systems like, e.g., the tetrapyrrole chains in phytochrome. First, for accurately computing the chromophore spectra, one will need a more efficient and more accurate DFT module.<sup>51</sup> Second, for the speedy and repeated computation of the polarization in an extended neighborhood of the chromophore, one may resort to an even faster though less accurate approach such as the self-consistent charge density functional tight-binding method.<sup>52</sup> Of course, instead of such a polarized force field, the use of a reliable polarizable force field for proteins would be desirable. However, such force fields are not yet established. Nevertheless, within a MM–MD scenario involving polarized force fields, a prediction of the L intermediate in the BR photocycle through microsecond simulations seems to be technically feasible, as is

demonstrated by the recent 10  $\mu$ s simulation of a small protein in solution.<sup>53</sup> As a result, there is hope that the mechanism converting BR light energy into proton pumping can be revealed.

**Acknowledgment.** The authors gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (SFB 533, projects C1 and C3). The authors thank B. Schropp for technical support in the DFT/MM calculations.

**Supporting Information Available:** On 14 pages, the SI provides 11 figures and associated text explaining and discussing the additional material. The SI visualizes the C–C stretching modes of an RSBH<sup>+</sup> (Figures 17, 18), illustrates the effects of the DFT fragment size, of thermal fluctuations, and of conformational transitions on the chromophore's structure and IR spectra (Figures 19–22), discusses the spectral effects of water polarization (Figures 23, 24) and of various isotope substitutions (Figure 25), and explains the structural and spectral consequences of altering the hydrogen bonds in the chromophore binding pocket (Figures 26–29). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Babitzki, G.; Denschlag, R.; Tavan, P. *J. Phys. Chem. B*. DOI: 10.1021/jp902428x.
- (2) Lanyi, J. K. *Annu. Rev. Physiol.* **2004**, *66*, 665–688.
- (3) Kühlbrandt, W. *Nature* **2000**, *406*, 569–570.
- (4) Schulten, K.; Tavan, P. *Nature* **1978**, *272*, 85–86.
- (5) Schulten, K.; Schulten, Z.; Tavan, P. An Isomerization Model for the Pump Cycle of Bacteriorhodopsin. In *Information and Energy Transduction in Biological Membranes*; Bolis, L., Helmreich, E. J. M., Passow, H., Eds.; Allan R. Liss, Inc.: New York, 1984.
- (6) Smith, S. O.; Myers, A.; Pardo, J.; Winkel, C.; Mulder, P.; Lugtenburg, J.; Mathies, R. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 2055–2059.
- (7) Alshuth, T. Kinetische und strukturelle Untersuchungen am Chromophor von Bacteriorhodopsin mit Hilfe zeitaufgelöster Resonanz-Raman-Spektroskopie. Doktorarbeit, Georg-August Universität Göttingen, 1985.
- (8) Gerwert, K. Transduktion der Lichtenergie in Protonen-Transfer-Reaktionen beim Bacteriorhodopsin: Eine Untersuchung mit Hilfe der zeitaufgelösten IR- und statischen FTIR-Differenzspektroskopie. Doktorarbeit, Albert-Ludwigs-Universität Freiburg im Breisgau, 1985.
- (9) Gerwert, K.; Siebert, F. *EMBO J.* **1986**, *5*, 805–811.
- (10) Fodor, S. P. A.; Pollard, W. T.; Gebhard, R.; van den Berg, E. M. M.; Lugtenburg, J.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2156–2160.
- (11) Großjean, M. F.; Tavan, P.; Schulten, K. *Eur. Biophys. J.* **1989**, *16*, 341–349.
- (12) Fahmy, K.; Großjean, M. F.; Siebert, F.; Tavan, P. *J. Mol. Struct.* **1989**, *214*, 257–288.
- (13) Hage, W.; Kim, M.; Frei, H.; Mathies, R. A. *J. Phys. Chem.* **1996**, *100*, 16026–16033.
- (14) Rödig, C.; Chizhov, I.; Weidlich, O.; Siebert, F. *Biophys. J.* **1999**, *76*, 2687–2701.
- (15) Lanyi, J. K. *Biochim. Biophys. Acta* **2004**, *1658*, 14–22.
- (16) Lanyi, J. K.; Schobert, B. *J. Mol. Biol.* **2007**, *365*, 1379–1392.
- (17) Braiman, M.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 403–407.
- (18) Bondar, A.-N.; Suhai, S.; Fischer, S.; Smith, J. C.; Elstner, M. *J. Struct. Biol.* **2007**, *157*, 454–469.
- (19) Royant, A.; Edman, K.; Ursby, T.; Pebay-Peyroula, E.; Landau, E. M.; Neutze, R. *Nature* **2000**, *406*, 569–570.
- (20) Lanyi, J. K.; Schobert, B. *J. Mol. Biol.* **2003**, *328*, 439–450.
- (21) Edman, K.; Royant, A.; Larsson, G.; Jacobson, F.; Taylor, T.; Van Der, S. D.; Landau, E. M.; Pebay-Peyroula, E.; Neutze, R. *J. Biol. Chem.* **2004**, *279*, 2147–2158.
- (22) Kouyama, T.; Nishikawa, T.; Tokuhisa, T.; Okumura, H. *J. Mol. Biol.* **2004**, *335*, 531–546.
- (23) Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 883–888.
- (24) Fahmy, K.; Siebert, F.; Tavan, P. *Biophys. J.* **1991**, *60*, 989–1001.
- (25) Lórenz-Fonfría, V. A.; Furutani, Y.; Kandori, H. *Biochemistry* **2008**, *47*, 4071–4081.
- (26) Dioumaev, A. K.; Lanyi, J. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9621–9626.



- (27) For physiological temperatures, the kinetic competition of the  $L \rightarrow BR_{568}$  shunt reaction with the kinetically preferred deprotonation of the  $RSBH^+$  marking the  $L \rightarrow M_1$  transition has been first discussed in ref 5.
- (28) Bondar, A.-N.; Elstner, M.; Suhai, S.; Smith, J. C.; Fischer, S. *Structure* **2004**, *12*, 1281–1288.
- (29) Braun-Sand, S.; Sharma, P. K.; Chu, Z. T.; Pislakov, A. V.; Warshel, A. *Biochim. Biophys. Acta* **2008**, *1777*, 441–452.
- (30) Diller, R.; Stockburger, M. *Biochemistry* **1988**, *27*, 7641–7651.
- (31) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452–10467.
- (32) The case of an unprotonated RSB featuring a polyene-like pattern of alternating single and double bonds is much simpler. Here, even an empirical modeling of the force field may be feasible. See, e.g., ref 54.
- (33) Curry, B.; Broek, A.; Lugtenburg, J.; Mathies, R. A. *J. Am. Chem. Soc.* **1982**, *104*, 5274–5286.
- (34) Smith, S. O.; Myers, A.; Mathies, R.; Pardo, J.; Winkel, C.; van den Berg, E.; Lugtenburg, J. *Biophys. J.* **1985**, *47*, 653–664.
- (35) Smith, S. O.; Braiman, M. S.; Myers, A. B.; Pardo, J. A.; Courtin, J. M. L.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. *J. Am. Chem. Soc.* **1987**, *109*, 3108–3125.
- (36) Großjean, M. F.; Tavan, P.; Schulten, K. *J. Phys. Chem.* **1990**, *94*, 8059–8069.
- (37) Schmitz, M.; Tavan, P. On the art of computing the IR spectra of molecules in condensed phase. In *Modern methods for theoretical physical chemistry of biopolymers*; Starikov, E. B., Tanaka, S., Lewis, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2006.
- (38) Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899–911.
- (39) Hutter, J.; Alavi, A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. “CPMD V3.9”, Copyright IBM Corp and MPI für Festkörperforschung Stuttgart, 2004; see [www.cpmid.org](http://www.cpmid.org).
- (40) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (41) Perdew, J.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- (42) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2005.
- (43) Nonella, M.; Mathias, G.; Tavan, P. *J. Phys. Chem. A* **2003**, *107*, 8638–8647.
- (44) Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, *107*, 316–322.
- (45) Großjean, M. F.; Tavan, P. *J. Chem. Phys.* **1988**, *88*, 4884–4896.
- (46) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (47) MacKerell, A.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (48) Lansing, J. C.; Hohwy, M.; Jaroniec, C. P.; Creemers, A. F. L.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **2002**, *41*, 431–438.
- (49) Alshuth, T.; Stockburger, M. *Photochem. Photobiol.* **1986**, *43*, 55–66.
- (50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (51) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (52) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (53) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75–L77.
- (54) Ames, J. B.; Fodor, S. P. A.; Gebhard, R.; Raap, J.; van den Berg, E. M. M.; Lugtenburg, J.; Mathies, R. A. *Biochemistry* **1989**, *28*, 3681–3687.

JP902432E

Supplementary material to the manuscript

The infra-red spectra of the retinal chromophore in  
bacteriorhodopsin  
calculated by a DFT/MM approach

G. Babitzki and P. Tavan\*

Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität,  
Oettingenstr. 67, 80538 München, Germany

\*corresponding author, email: [tavan@physik.uni-muenchen.de](mailto:tavan@physik.uni-muenchen.de),  
phone: +49-89-2180-9220, fax: +49-89-2180-9202



## Normal modes of the BR chromophore

Figures 17 and 18 illustrate for the prominent double and single bond stretching bands in the IR spectrum of BR<sub>568</sub> the typical shapes of the associated normal modes as determined by our DFT/MM calculations. More specifically, the depicted data have been taken from a normal mode analysis of the structure 1C3W/MD using the PBR force field for the protein and the large DFT fragment **L**. For the double bonds the associated frequencies are given in column **L** of Fig. 10 (paper). However, because experience has shown that the compositions of the selected modes do not change much, e.g., upon thermal fluctuations of the structure we have assigned the normal modes to the experimental band positions in the figures.

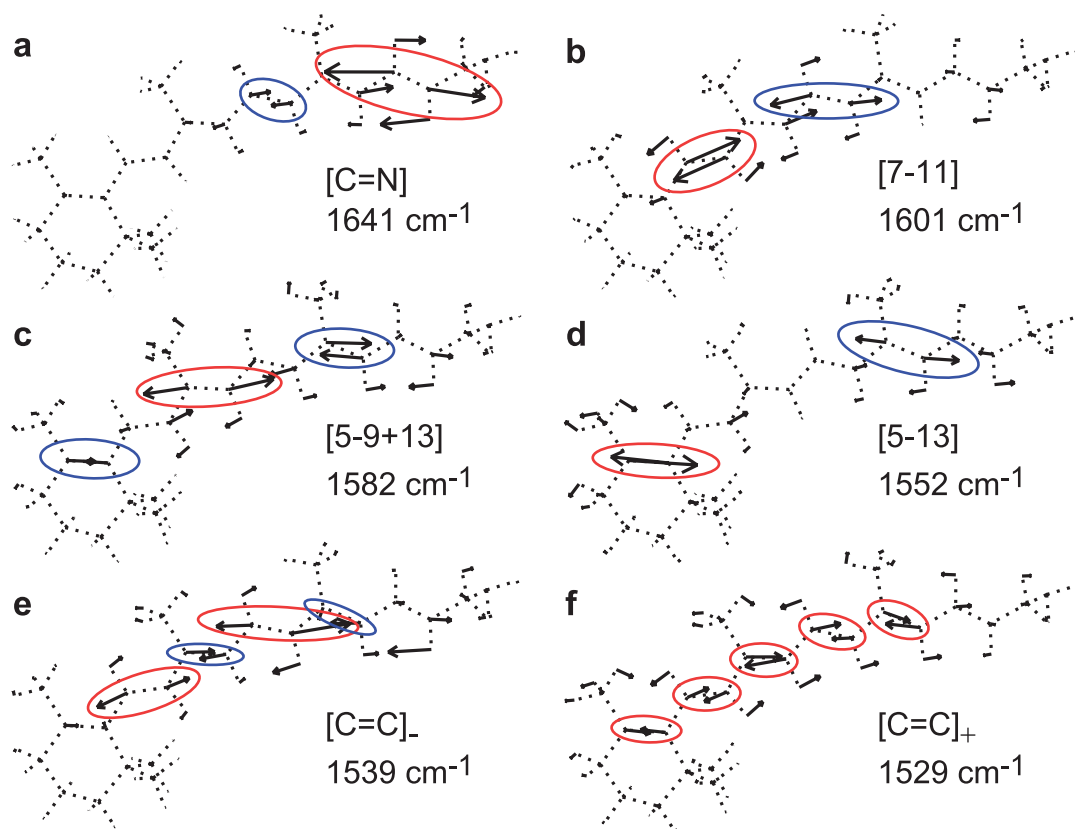


Figure 17: Graphical representations of the six C=C and C=N normal modes in BR<sub>568</sub> and the associated symbols characterizing the mode compositions (see the explanations in connection with Fig. 7). Different phases of local stretches are emphasized by color coding. The listed wave numbers are the experimental band positions<sup>1</sup> assigned by us to the depicted normal modes. The data employed for this graph were obtained from a DFT/MM normal mode analysis of structure 1C3W/MD with the PBR force field and DFT fragment **L**.

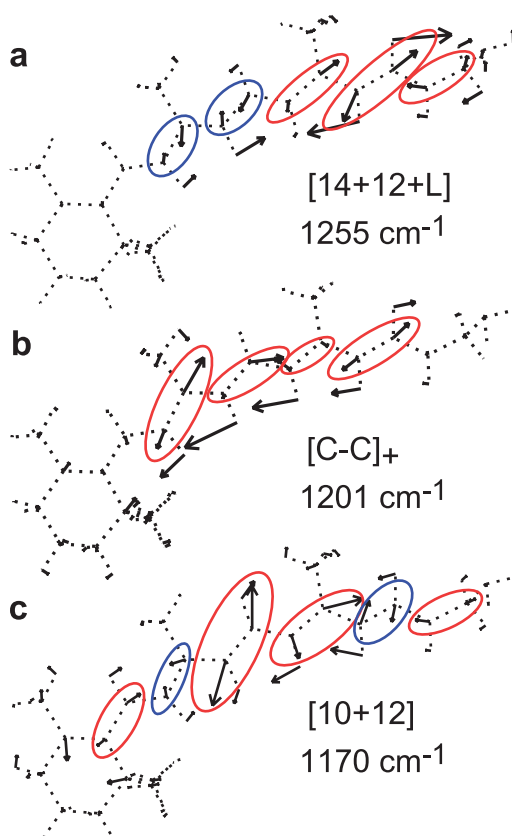


Figure 18: Graphical representations of the normal modes involving C—C single bond stretches assigned by us to the three intense fingerprint bands in the IR spectrum of BR<sub>568</sub>. The listed wavenumbers are the experimental band positions given by Gerwert.<sup>2</sup> For further explanations see the caption to Fig. 17.

---

## Effect of DFT fragment on chromophore structure

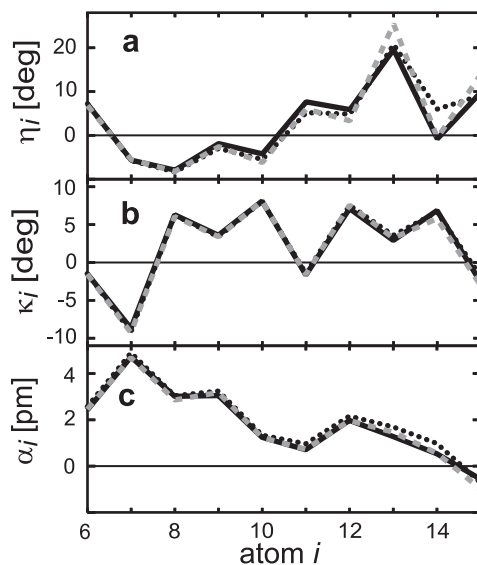


Figure 19: The effect of the DFT-fragment size (cf. Fig. 2 in the paper) on the structure of the  $\text{RSBH}^+$  (see the caption to Fig. 4 in the paper for explanations of the shown observables). The shown DFT/MM data were obtained for the 100 K structure 1C3W/MD using the PBR force field for the apoprotein. Black dotted: fragment **S**. Gray dashed: fragment **M**. Black solid: fragment **L**. The displayed data are discussed in the text of the paper near Fig. 9.

---

## Structural effects of thermal fluctuations

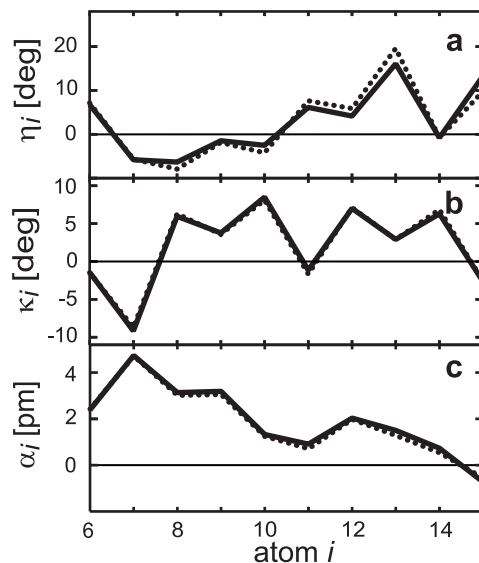


Figure 20: For conformation  $C_1$  the average chromophore structure at 300 K (solid) is compared with the structure of a 100 K snapshot (dotted) using the DFT/MM partitioning  $\mathbf{L}$  in both cases (see the caption to Fig. 4 for explanations of the shown observables and the text near Fig. 11 for further explanations).

As displayed by Fig. 20a, for conformation  $C_1$  the inclusion of room temperature fluctuations leads to somewhat smaller torsional angles in most parts of the chromophore (with the exception of the C=N group). According to Fig. 20c the alternation slightly increases with the temperature (from an average value of 1.8 pm at 100 K to 1.9 pm at 300 K). However, the curvature of the chromophore (Fig. 20b) is nearly invariant upon change of temperature.

## Effects of conformational transitions

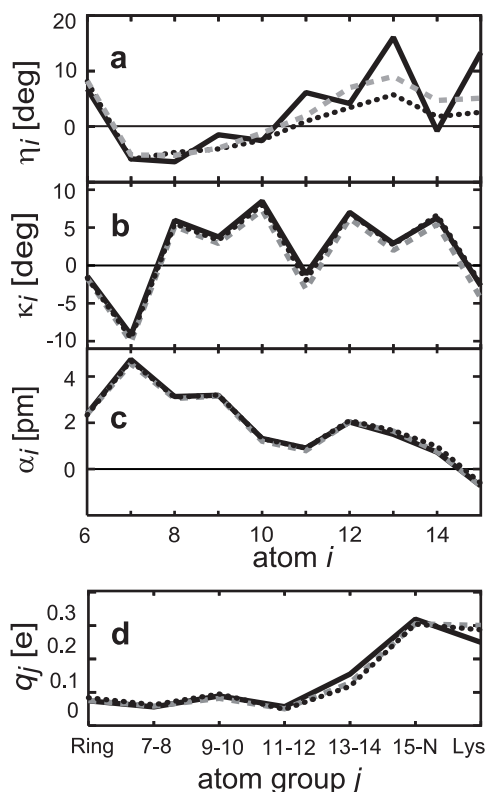


Figure 21: Average chromophore helicities (a), curvatures (b), bond alternations (c), and charge distributions (d) in the 300 K conformational ensembles C<sub>1</sub> (black solid), C<sub>2</sub> (black dotted), and C<sub>3</sub> (gray dashed).

In part I of this work, the conformational space of the lysine chain connecting the chromophore with the protein backbone has been extensively sampled at 300 K using a particularly designed Hamiltonian replica exchange approach.<sup>3</sup> According to this sampling, at 300 K not only the crystallographic conformation C<sub>1</sub> is occupied but also at least one additional conformation C<sub>2</sub> and, possibly, though with a much smaller frequency, also a third conformation C<sub>3</sub>. From extended 300 K MD simulations sampling the thermal fluctuations within the three, very slowly interconverting conformational states, structural ensembles have been extracted covering 25 (C<sub>1</sub>) and 20 (C<sub>2</sub>, C<sub>3</sub>) snapshots, respectively.

Fig. 21 gives an overview over the average chromophore structures within the the three conformational substates. Quite apparently there are distinct influences of the conformational transitions on a few, and small effects on most structural observables. According to Fig. 21a the helicities are strongly affected by conformational transitions. Whereas in C<sub>1</sub> the overall helicity  $\eta_{\text{pol}}$  of the polyene moiety is large (23°), it is smaller by about a factor 0.5 for C<sub>3</sub> and nearly vanishes for C<sub>2</sub>. Thus in conformations C<sub>2</sub> and C<sub>3</sub> the steric strain, which acts in C<sub>1</sub> and causes in this conformation substantial deviations from planarity, is strongly reduced. The seemingly small differences in curvature visible in Fig. 21b add up to a substantial difference

between conformations  $C_1$  and  $C_2$ , on the one hand, and conformation  $C_3$  on the other. The former states are characterized by average total curvatures  $\kappa_{\text{tot}} \approx 20^\circ$  signifying a considerable curving strain exerted by the protein. Conformation  $C_3$ , in contrast, has an average total curvature  $\kappa_{\text{tot}} \approx 10^\circ$ . Thus, with respect to the curvature, conformation  $C_3$  resembles the isolated chromophore VAC.

As is hardly visible in Fig. 21c, the average bond alternation increases in the sequence  $C_3$  (1.88 pm)  $\rightarrow$   $C_1$  (1.92 pm)  $\rightarrow$   $C_2$  (1.98 pm). Correspondingly and also barely detectable in the curves of Fig. 21d, the localization of positive charge  $q_{\text{SB}}$  near the Schiff base increases also a little in the same sequence starting 0.71e ( $C_1$ ) with increments of 0.01e. From the latter two observables one thus expects blueshifts of the C=C double bond stretches and redshifts of the C—C single bond stretches in the sequence  $C_3 \rightarrow C_1 \rightarrow C_2$ .

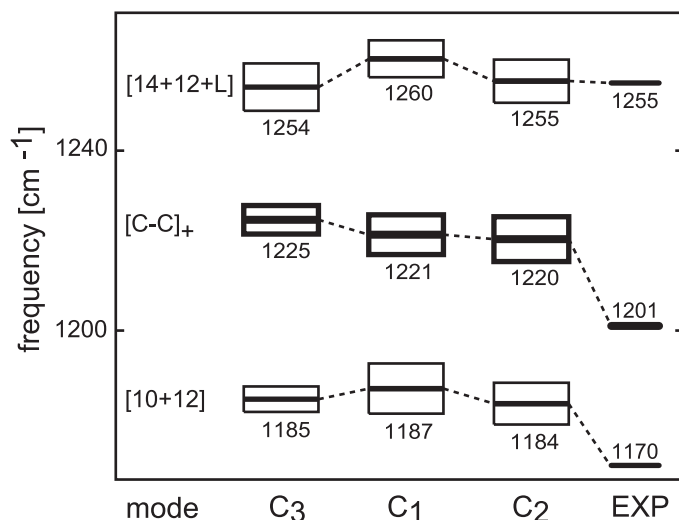


Figure 22: Effect of the respective conformational state  $C_i$ ,  $i = 1, 2, 3$ , on the average frequencies and band widths of the three fingerprint C—C single bond stretches. Experimental data are given as a reference (cf. Figs. 7 and 8 in the paper).

Fig. 22 shows that the expected redshifts are found solely for the most intense  $[C—C]_+$  mode. The frequency of this strongly delocalized C—C stretching vibration thus seems to be mainly determined by the size of the charge delocalization whereas the frequencies of the other modes are also affected by differences of geometry. Like Fig. 8 in the paper also Fig. 22 reveals that our DFT/MM approach sizably overestimates the frequencies of the fingerprint bands. Here, the root mean square deviations from the experimental data are 16.1 cm $^{-1}$  ( $C_3$ ), 15.7 cm $^{-1}$  ( $C_1$ ), and 13.7 cm $^{-1}$  ( $C_2$ ) and thus about as large as the 15.0 cm $^{-1}$  characterizing column PBR in Fig. 8 (paper). Furthermore, in the C—C stretching range the IR spectra of the three conformational substates are predicted to be very similar because the band widths calculated for the fingerprint bands are large and because the associated frequency differences are small.

## Effect of polarized water models on spectra

As outlined in part I of this work we employed the non-polarized TIP3P water model for the crystallographic water molecules in the chromophore binding pocket because we wanted to check whether the polarization of the protein material suffices to stabilize the pentagonal hydrogen bonded network in the center of the purple membrane. This turned out to be the case. However, it came as a surprise that the dipole moments of the tightly bound water molecules were predicted by DFT/MM to be larger than dipole moments of water molecules in a bulk aqueous phase. Because this insight came after we had executed all the MD simulations on BR, we have no access to equilibrium ensembles with more polar water molecules. Therefore we used the standard TIP3P water models in all our DFT/MM calculations of the chromophore vibrational spectra with the exception of the result documented below.

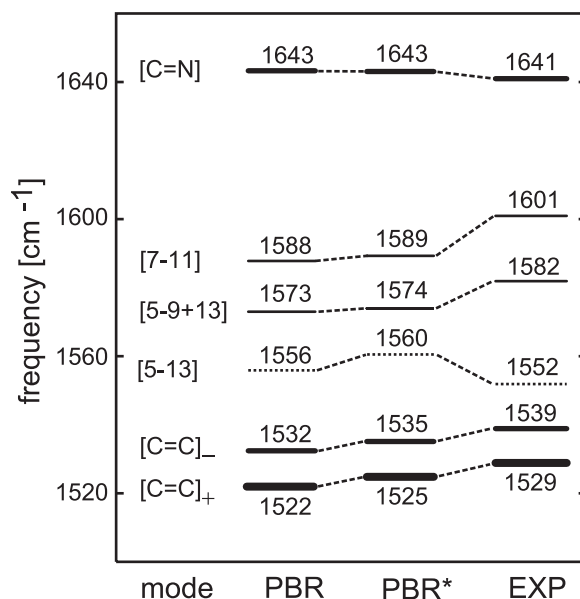


Figure 23: Double bond stretching frequencies calculated for the large DFT-fragment **L** and the cryogenic all-atom model 1C3W/MD using the polarized force field PBR in combination with TIP3P water ("PBR") or using PBR combined with more polar water models ("PBR\*") derived from TIP3P by increasing the dipole moment from 2.35 D to 2.59 D. Part I of this work has presented DFT/MM calculations on the three tightly bound water molecules in the chromophore binding pocket rendering this surprisingly enhanced dipole moment. Experimental frequencies are given as a reference.

The calculation documented in Fig. 23 compares the spectral effect of exchanging TIP3P water (column PBR) with more polar and, thus, more tightly bound water molecules. The data in column PBR are copies of the data in Fig. 10, column **L** in the main text. We have verified that the underlying structures resulting from DFT/MM minimizations of the starting structure 1C3W/MD are essentially identical. Thus, the spectral changes have to be exclusively attributed to the enhance dipole moments of the three water molecules in the chromophore binding pocket.

According to Fig. 23 the C=C stretching frequencies are almost uniformly shifted by  $2 \text{ cm}^{-1}$

to higher frequencies upon the use of more polar water models. The C=N frequency, in contrast, remains essentially invariant. As a result the root mean square deviation of the calculated frequencies from the experimental values reduces to  $7.2\text{ cm}^{-1}$  (PBR\*) from originally  $7.8\text{ cm}^{-1}$  (PBR). It is pleasing that an improved modeling of the electrostatics in the binding pocket moves the computational results closer to the observations.

The blue-shift of the C=C frequencies argues for a reduced charge delocalization in the PBR\* model. In fact, inspecting the corresponding data one finds that the charge  $q_{SB}$  in the Schiff base region increases by 6 % with more strongly polar water molecules. Correspondingly, also the average bond alternation increases by 11%. Thus, the use of more polar water models can diminish the key remaining drawback of our DFT/MM models for the BR chromophore which is the overestimated charge delocalization.

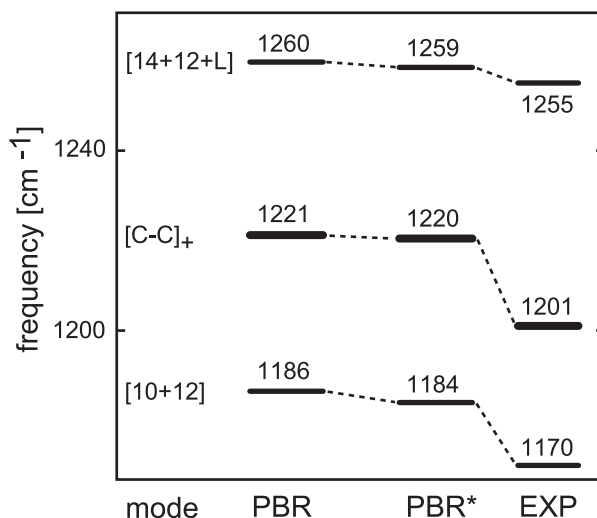


Figure 24: Influence of more polar water models on the single bond stretching frequencies of the fingerprint bands. See the Caption to Fig. 23 for explanation.

As documented by Fig. 24, the slightly larger localization of the positive charge near the Schiff base shows up in the fingerprint region of the C—C single bond stretches as a nearly uniform red-shift of the frequencies by  $1\text{-}2\text{ cm}^{-1}$ . Thus, also here the inclusion of more polar water models brings the predictions a little closer to the observations although sizable overestimates of the charge delocalization and, correspondingly, of the single bond stretching force constants remain.



## Selected isotope effects on the room temperature IR spectra

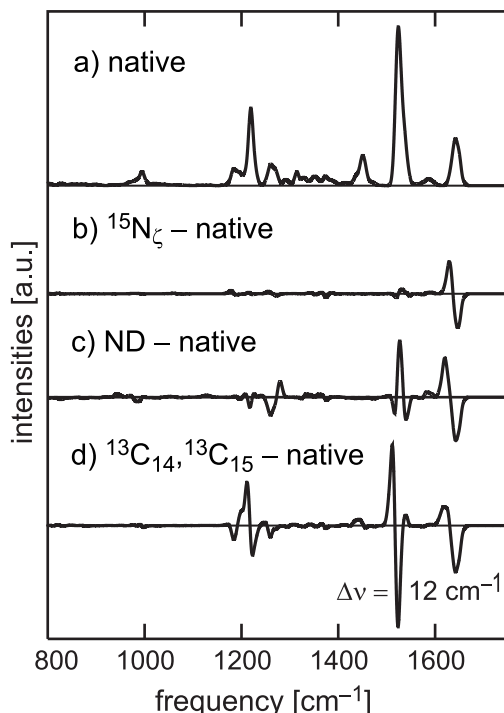


Figure 25: Absolute and difference IR spectra from DFT/MM-INMA expressing isotope effects: a) absolute spectrum of the unlabeled BR chromophore, copied from Fig. 14b in the paper. Difference spectra for b)  $^{15}\text{N}_\zeta$ , c) ND, and d)  $^{13}\text{C}_{14}, ^{13}\text{C}_{15}$  labels.

As an example, consider the composition of the dominant ethylenic stretching mode  $[\text{C}=\text{C}]_+$ . The empirical calculations assigned a predominant  $\text{C}_{11}=\text{C}_{12}$  stretching character to this mode thus overestimating the red-shifts by a factor of two which are caused by  $^{13}\text{C}_{11}$  or  $^{13}\text{C}_{12}$  isotope labels.<sup>4</sup> As is immediately apparent from an inspection of Fig. 17, the  $[\text{C}=\text{C}]_+$  mode resulting from DFT/MM descriptions is much more delocalized. This delocalization is compatible with the observation that a  $^{13}\text{C}$  label at any position between  $\text{C}_9$  and  $\text{C}_{14}$  causes approximately the same  $7\text{ cm}^{-1}$  red-shift of the dominant ethylenic band.<sup>4</sup> For instance, introducing a  $^{13}\text{C}_{14}, ^{13}\text{C}_{15}$  double label yields experimentally a  $10\text{ cm}^{-1}$  red-shift of the  $[\text{C}=\text{C}]_+$  band<sup>4</sup> whereas we calculate according to Fig. 25d a value of  $12\text{ cm}^{-1}$ .

## Spectral effect of changing the couterion

In part I of this work the MD simulations on BR with the C22 force field yielded an artificial conformation which is denoted as  $C_{nn}$  and features a hydrogen bond with the Schiff base. In part II we compare the INMA frequencies of the single and double bond stretches obtained for the  $C_{nn}$  conformation and for the C22 force field with INMA frequencies derived for the room temperature conformational mixture  $0.58 C_1 + 0.42 C_2$  and for the force field PBR. This comparison comprises the changes of (i) the conformational ensemble and of (ii) the force field parameterization. If one wants to identify the spectral effects of point (i), i.e. of the altered ensemble, one must be able to estimate the effects of point (ii), i.e. of the altered force field.

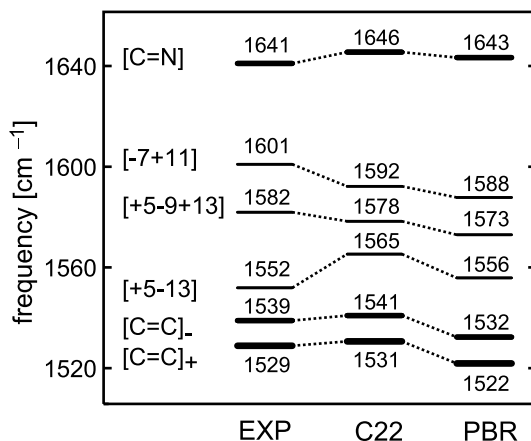


Figure 26: Effect of force field (PBR vs. C22) on the C=C and C=N stretch frequencies of the BR chromophore from DFT/MM computations on the structure 1C3W/MD using the partition **L**. For reference also the experimental frequencies are given.

Fig. 26 illustrates the typical sizes and directions of the spectral shifts induced by by using C22 instead of PBR as the force field for the apoprotein. Replacing the polarized force field PBR by C22 apparently shifts the double bond stretching frequencies in a nearly homogeneous fashion by  $6 \text{ cm}^{-1}$  to the blue. In these calculations the structure 1C3W/MD and the DFT/MM partition **L** were used. Note that a similar comparison is displayed by Fig. 7. It differs from the one given above by choosing the DFT/MM partition **M** instead of **L**. For **M** the replacement of PBR by C22 entails a uniform  $7 \text{ cm}^{-1}$  blue-shift of the double bond stretching frequencies. As a result one would expect also for INMA spectra calculated from identical conformational ensembles a similar blue-shift of the double bond stretching frequencies upon replacing PBR by C22.

In part I of this work we have characterized the various structural ensembles obtained by MD simulations by two conformational coordinates  $d_1$  and  $d_2$ . These coordinates measure the distances between the Schiff base proton and the oxygen of the nearest water molecule ( $d_1$ ) or the nearest oxgen of the carboxy group of Asp85 ( $d_2$ ), respectively. Free energy landscapes  $\Delta G(d_1, d_2)$  were derived from extended 300 K simulations and discussed.

Fig. 27a shows the free energy landscapes  $\Delta G(d_1, d_2)$  for a room temperature conformational mixture  $0.58 C_1 + 0.42 C_2$  obtained with the polarized force field PBR. The snapshots used for the INMA calculation on the PBR ensemble are marked as black dots in the  $(d_1, d_2)$ -plane.

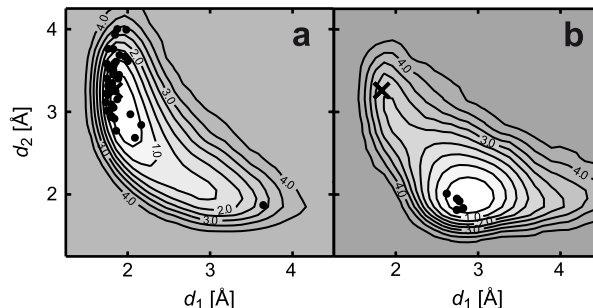


Figure 27: Free energy landscape  $\Delta G(d_1, d_2)$  (in kcal/mol) for (a) the PBR force field and the conformational mixture  $0.52 C_1 + 0.42 C_2$  and for (b) the C22 force field and conformation  $C_{nn}$  obtained in part I by extended MD simulations at 300 K. Also shown (dots) are the snapshots drawn from these trajectories and used for the INMA computation of IR spectra.

58 % of the 43 snapshots belong to conformation  $C_1$  and the remainder to  $C_2$ . Fig. 27b is a corresponding representation of the 300 K free energy landscape obtained for conformation  $C_{nn}$  from MD simulations with the C22 force field. Here, only 5 snapshots were used for the computation of an IR spectrum with the INMA procedure.

According to Fig. 27a nearly all PBR-INMA snapshots exhibit the experimental hydrogen bond of the Schiff base with a water molecule. Only a single snapshot belongs to the unstable hydrogen bond with Asp85. In contrast, according to Fig. 27b all C22-INMA snapshots exhibit an artificial hydrogen bond with Asp85.

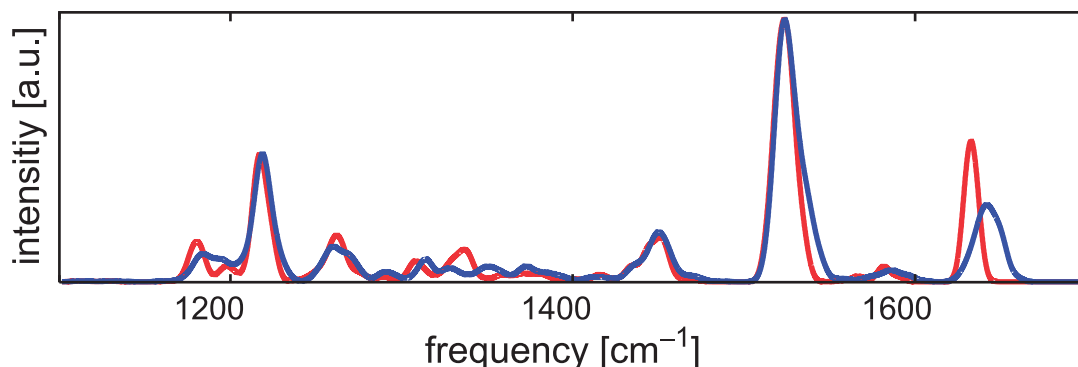


Figure 28: IR spectra computed by INMA from 300 K MD trajectories employing the PBR (blue) and C22 (red) force fields. The underlying snapshot ensembles are characterized in Fig. 27.

Fig. 28 compares the IR spectra computed by INMA from the PBR and C22 snapshot ensembles introduced above. The spectra have been constructed using the procedures explained in the text discussing Fig. 14.

Contrary to the blue shift expected from Figs. 7 and 26 upon using C22 instead of PBR partial charges, in the  $C_{nn}$  conformation of the C22 ensemble the bands associated with double bond stretches are shifted to the red. As displayed by Figs. 16 and 28, this red shift is most

pronounced for the high frequency C=N stretch (at  $1642\text{ cm}^{-1}$  for PBR). Thus, point (i), i.e. the change of the structural ensemble apparently can overcompensate the  $6\text{ cm}^{-1}$  blue shift expected from point (ii), i.e. from replacing PBR by C22.

Interestingly, in the  $C_{nn}$  conformation of the C22 ensemble also the frequencies of the single bond stretches are shifted by about  $2\text{ cm}$  to the red (data not shown). A uniform red-shift of all C-C stretching frequencies was not observed previously in any of those calculations which left the chromophore geometry fixed while changing other parameters. Here, a red-shift of double bond stretches was always accompanied by a blue shift of single bond stretches and vice versa. Thus the surprising red shift of all C-C stretching frequencies must be caused by an altered chromophore structure in the  $C_{nn}$  conformation.

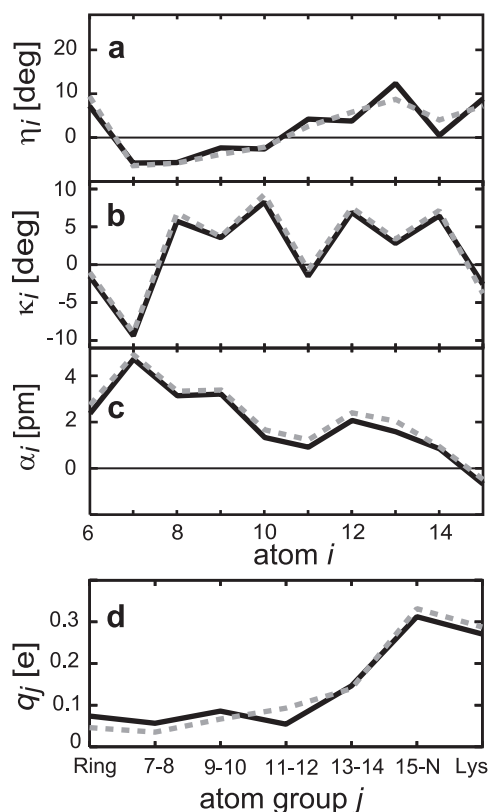


Figure 29: Average chromophore helicities (a), curvatures (b), bond alternations (c), and charge distributions (d) at 300 K in the native conformational mixture  $0.52 C_1 + 0.42 C_2$  obtained with PBR (black solid) and in the  $C_{nn}$  conformation of the C22 ensemble (gray dashed).

Fig. 29 compares the average 300 K structure of the  $RSBH^+$  obtained for the native conformational mixture in the PBR ensemble (black solid) with the one calculated for the  $C_{nn}$  conformation in the C22 ensemble (gray dashed). According to Fig. 29b, the chromophore curvature is a little smaller in the native conformation than in  $C_{nn}$ . Fig. 29a reveals sizable differences of the dihedral angles in the Schiff base region of the chromophore. In  $C_{nn}$  the twists around the  $C_{12}-C_{13}$  and  $C_{14}-C_{15}$  single bonds increase by 2 deg and 4 deg, respectively, whereas the twists around the  $C_{13}=C_{14}$  and  $C_{15}=N$  double bonds decrease by 4 deg and 2 deg. Furthermore, in  $C_{nn}$  the charge is more strongly localized near the Schiff base as is signified by

---

an  $0.1e$  increase of  $q_{\text{SB}}$ . In line with the increased charge localization also the alternation is larger for  $C_{nn}$ .

For an unchanged geometry the noted changes of the charge and the alternation would argue for a red-shift of the single bond stretches and a blue shift of the double bond stretches. The calculated red-shifts of the double bond stretches thus must be caused by the different twisting patterns in the Schiff base region of the  $\text{RSBH}^+$  strongly overcompensating the effects of enhanced bond alternation and charge localization. We conclude that the vibrational spectra of an  $\text{RSBH}^+$  chromophore are extremely sensitive probes not only for external fields but also for fine details of its geometry.

## References

1. Alshuth, T.; Stockburger, M. *Photochem. Photobiol.* **1986**, *43*, 55-66.
2. Gerwert, K. *Transduktion der Lichtenergie in Protonen-Transfer-Reaktionen beim Bacteriorhodopsin: Eine Untersuchung mit Hilfe der zeitaufgelösten IR- und statischen FTIR-Differenzspektroskopie*, Doktorarbeit, Albert-Ludwigs-Universität Freiburg im Breisgau, 1985.
3. Babitzki, G.; Denschlag, R.; Tavan, P. *J. Phys. Chem. B* **2008**, submitted.
4. Smith, S. O.; Braiman, M. S.; Myers, A. B.; Pardo, J. A.; Courtin, J. M. L.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. *J. Am. Chem. Soc.* **1987**, *109*, 3108-3125.

### 3 Untersuchung der Konformationsdynamik eines $\beta$ -Hairpinpeptides

Wie ich in der Einleitung betont habe, ist das bisher betrachtete Protein BR zu komplex, um mit den gegenwärtig verfügbaren Rechenkapazitäten eine vollständige Simulationsbeschreibung der lichtinduzierten Konformationsdynamik zu gestatten. Einfacher ist demgegenüber das dort eingeführte  $\beta$ -Hairpinpeptid. Die nachfolgend geschilderten Untersuchungen zur Konformationsdynamik des  $\beta$ -Hairpinpeptides kombinieren zeitaufgelöste IR-Spektroskopie mit MM-MD Simulationen der lichtinduzierten Dynamik und DFT-MM Berechnungen der IR Spektren (AmidI-Bereich) des Hairpins in Methanol. Diese DFT/MM Rechnungen stellen meinen Hauptbeitrag zum Artikel

Tobias E. Schrader, Wolfgang J. Schreier, Thorben Cordes, Florian O. Koller, Galina Babitzki, Robert Denschlag, Christian Renner, Markus Löweneck, Shou-Liang Dong, Luis Moroder, Paul Tavan, and Wolfgang Zinth: „Light-triggered  $\beta$ -hairpin folding and unfolding“ *Proceedings of the National Academy of Sciences* 104, 15729-15734 (2007),

den ich gemeinsam mit Tobias E. Schrader, Wolfgang J. Schreier, Thorben Cordes, Florian O. Koller, Robert Denschlag, Christian Renner, Markus Löweneck, Shou-Liang Dong, Luis Moroder, Paul Tavan, und Wolfgang Zinth publiziert habe.

Zu dieser Veröffentlichung findet sich auf der Internetseite des *Proceedings of the National Academy of Sciences*<sup>1</sup> Zusatzmaterial, das im Anschluss an den Artikel hier ebenfalls abgedruckt ist. Dieses Material ist wichtig, weil nur dort die eingesetzten Rechenmethoden genauer beschrieben werden.

---

<sup>1</sup><http://www.pnas.org>





# Light-triggered $\beta$ -hairpin folding and unfolding

Tobias E. Schrader\*, Wolfgang J. Schreier\*, Thorben Cordes\*, Florian O. Koller\*, Galina Babitzki\*, Robert Denschlag\*, Christian Renner†‡, Markus Löweneck†§, Shou-Liang Dong†, Luis Moroder†, Paul Tavan\*, and Wolfgang Zinth\*¶

\*Lehrstuhl für BioMolekulare Optik and Munich Center for Integrated Protein Science, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, 80538 Munich, Germany; and †Max-Planck-Institut für Biochemie, Am Klopferspitz 18, 82152 Martinsried, Germany

Communicated by Joshua Jortner, Tel Aviv University, Tel Aviv, Israel, August 7, 2007 (received for review March 30, 2007)

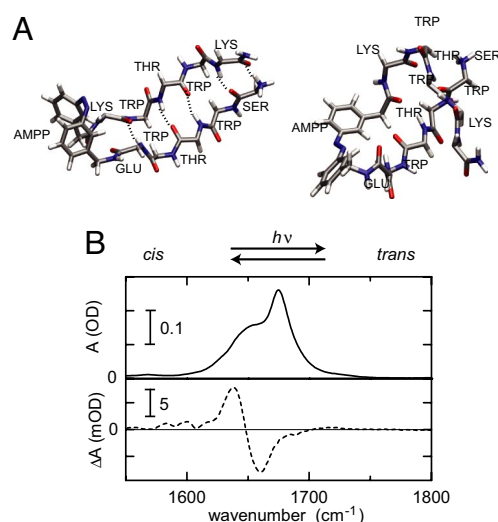
A light-switchable peptide is transformed with ultrashort pulses from a  $\beta$ -hairpin to an unfolded hydrophobic cluster and vice versa. The structural changes are monitored by mid-IR probing. Instantaneous normal mode analysis with a Hamiltonian combining density functional theory with molecular mechanics is used to interpret the absorption transients. Illumination of the  $\beta$ -hairpin state triggers an unfolding reaction that visits several intermediates and reaches the unfolded state within a few nanoseconds. In this unfolding reaction to the equilibrium hydrophobic cluster conformation, the system does not meet significant barriers on the free-energy surface. The reverse folding process takes much longer because it occurs on the time scale of 30  $\mu$ s. The folded state has a defined structure, and its formation requires an extended search for the correct hydrogen-bond pattern of the  $\beta$ -strand.

density functional theory calculation | peptide folding | TrpZip2 | ultrafast infrared spectroscopy

Folding is a key process during the formation of a functional protein after the synthesis of its amino acid chain (1). During folding, the amino acid chains are rearranged in highly complex processes, for which a detailed understanding is still missing. Straightforward solutions of the folding problem are prevented by the high dimensionality of the protein conformational spaces and by the wide range of relevant time scales. Thus, for complexity reduction, one has to focus on typical protein substructures, which are small enough to be analyzed by present-day technology. With corresponding peptide model systems there is a chance to monitor the formation of secondary structures, to identify characteristic intermediate states of these folding dynamics, and to carry out realistic simulations on a molecular level. An interesting class of model systems are light-triggered peptides, within which the incorporation of a photoresponsive element can initiate structural changes.

Among corresponding photoresponsive chromophores, azobenzene derivatives have become a popular choice when it comes to selecting a fast conformational trigger for peptide refolding, because their structure significantly changes on time scales of a few hundred femtoseconds after photoexcitation. Indeed, an azobenzene chromophore, used as a backbone element within cyclic peptides, was shown to induce large-scale conformational changes, whose dynamics could be monitored by visible (2) and IR (3) spectroscopy. Moreover, the experimental results clearly showed that the initial, strongly driven conformational changes of the peptide proceeded within a few picoseconds after the trigger event. Externally linked to an  $\alpha$ -helix, azobenzene also was used for unfolding (4) and refolding (5) of an  $\alpha$ -helical model peptide in the nanosecond to microsecond time range.

A prominent secondary structure motif is the  $\beta$ -sheet, for which a  $\beta$ -hairpin represents a minimal model (6, 7). We and others (8–10) have recently focused on the design of photocontrolled  $\beta$ -hairpin peptides. To enable an ultrafast initiation of  $\beta$ -hairpin folding and unfolding the azobenzene derivative 3-(3-aminomethylphenylazo)phenylacetic acid (AMPP) was used as a reversible light switch (8, 10), which, in the cis isomeric state, acts as  $\beta$ -turn mimetic (see Fig. 1A). Indeed, as shown by NMR, the peptide folds into  $\beta$ -hairpin conformations as long as the azo-



**Fig. 1.** Cis and trans conformations of the light-switchable peptide. (A) (Left) NMR-derived sample structure (10) of the  $\beta$ -hairpin peptide with the AMPP-chromophore in the cis conformation. (Right) The cis-to-trans photoisomerization of the chromophore drives the peptide into a different equilibrium conformational ensemble, for which a representative structure is shown. (B) IR absorption spectrum of the  $\beta$ -hairpin model peptide in the trans form (solid curve) and light-induced absorption difference observed upon illumination around 370 nm, leading to the formation of the cis form (broken curve).

benzene switch is in its cis form (10). This conformation is stabilized by hydrogen bonds between the two peptide  $\beta$ -strands and by hydrophobic interactions between aromatic residues incorporated into the two peptide strands. Upon cis-to-trans isomerization, the geometry of the switch prevents a likewise tight coupling between the N- and C-terminal parts of the peptide, and the peptide unfolds into an ensemble of globular structures, which are stabilized apparently by hydrophobic forces. Therefore we call the trans structures of the peptide “hydrophobic clusters.”

In this paper we present results obtained with this  $\beta$ -hairpin model by ultrafast UV pump, IR probe spectroscopy combined

Author contributions: C.R., L.M., P.T., and W.Z. designed research; T.E.S., W.J.S., T.C., F.O.K., G.B., and R.D. performed research; M.L. and S.-L.D. contributed new reagents/analytic tools; T.E.S., G.B., R.D., C.R., P.T., and W.Z. analyzed data; and T.E.S., W.J.S., T.C., F.O.K., G.B., R.D., C.R., L.M., P.T., and W.Z. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: AMPP, 3-(3-aminomethylphenylazo)phenylacetic acid; DFT, density functional theory; INMA, instantaneous normal mode analysis; MD, molecular dynamics; MM, molecular mechanics.

¶Present address: Deutsche Forschungsgemeinschaft, Kennedyallee 40, 53170 Bonn/Bad-Godesberg, Germany.

§Present address: Senn Chemicals AG, Guido Senn Strasse 1, CH-8157 Dielsdorf, Switzerland.

¶To whom correspondence should be addressed. E-mail: zinth@physik.uni-muenchen.de.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0707322104/DC1](http://www.pnas.org/cgi/content/full/0707322104/DC1).

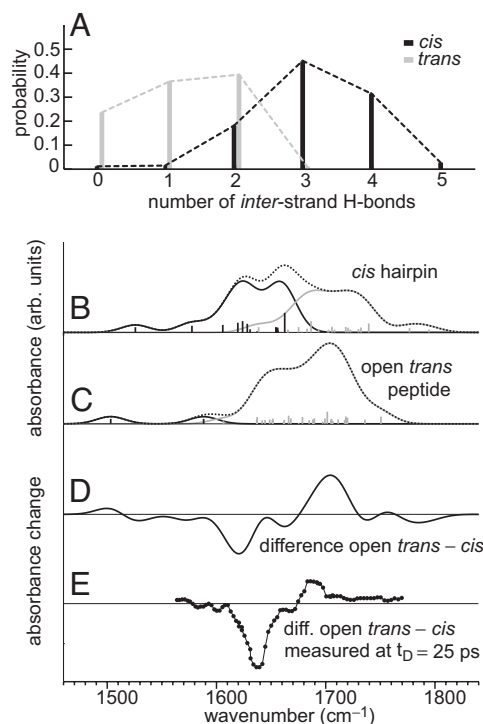
© 2007 by The National Academy of Sciences of the USA

with instantaneous normal mode analysis (INMA) and molecular dynamics (MD) simulations, which yielded detailed information not only on the kinetics but also on the structural intermediates of the light-triggered folding and unfolding processes.

**Stationary Spectroscopy of the Model Peptide.** Our IR experiments study the  $\beta$ -hairpin model dissolved in deuterated methanol. They focus on the spectral region of the amide I band generated predominantly by the C=O stretch vibrations of the backbone carbonyl groups. The amide I band is known to code the secondary structures of peptides and proteins (11, 12). A structural decoding of the absorption spectrum in the amide I region has to rely on empirical rules, because reliable theoretical approaches for interpretation of band shapes and peak positions in terms of protein structures are still lacking, although intense efforts are undertaken for improvement (13, 14). The empirical rules are based on experimental observations of model compounds (e.g., *N*-methylacetamide) and proteins (11, 12). For *N*-methylacetamide in a solvent with a low dielectric constant, the C=O group vibrates at high frequencies  $\nu/c \approx 1,680\text{--}1,700\text{ cm}^{-1}$  (14, 15). The vibrational frequency is lowered with rising polarity of the surroundings (14, 15) such that, e.g., in methanol, which is the mildly polar solvent used in our experiments, the C=O frequency is found at  $\approx 1,660\text{ cm}^{-1}$ . In water, which is a much more polar and protic solvent than methanol, the vibrational frequency is further reduced to  $\approx 1,625\text{ cm}^{-1}$  (14, 15). Additional changes of the spectral positions are found in proteins with dominant secondary structure elements. Here, the spectral positions are related to the structural arrangement of the polypeptide chains and the correspondingly shaped electrostatic and transition dipole interactions (11, 12).

The IR absorption spectrum of the model peptide in deuterated methanol with the azobenzene chromophore in the trans form is drawn in Fig. 1B (solid curve). This spectrum features a strong peak related to the amide I mode at  $1,675\text{ cm}^{-1}$ . A shoulder extends to lower frequencies down to the  $1,620\text{-cm}^{-1}$  range. Illumination of the sample around  $370\text{ nm}$  leads to the formation of *cis*-azobenzene and causes changes in the IR spectrum (see Fig. 1B, dashed curve). There is a pronounced increase in the IR absorption at  $1,640\text{ cm}^{-1}$ , whereas some decrease is found at  $1,660\text{ cm}^{-1}$ . Two weak bands of increased absorption at  $1,583$  and  $1,600\text{ cm}^{-1}$  are related to the azobenzene chromophore in the *cis* form. Assuming for a preliminary interpretation of the absorbance properties that the amide I band shape of our peptide model is mainly steered by local hydrogen bonding interactions and by the dielectric shielding of the amide dipoles caused by the mildly polar solvent (and thus neglecting the more complicated transition dipole interactions and corresponding secondary structure effects) yields the following picture, which agrees well with the NMR structural data (10). In the trans peptide, only few amide groups exhibit significant intramolecular hydrogen bonding, as is indicated by the weakness of the low-frequency shoulder at  $1,640\text{ cm}^{-1}$ . The pronounced higher-frequency peak at  $1,675\text{ cm}^{-1}$  suggests that most amide dipoles are only weakly shielded by the mildly polar solvent methanol (note here that, in water, one would expect a different behavior because of its larger polarity). Upon formation of the *cis* isomer of the azobenzene chromophore, the attached model peptide is transformed to the  $\beta$ -hairpin-like structure with an increased amount of strong intramolecular hydrogen bonds (10). This NMR result nicely fits the observed absorption changes featuring a pronounced absorption increase at  $1,640\text{ cm}^{-1}$  (see Fig. 1B, dashed curve).

**Computational Description of the Amide I Spectral Changes.** For the given, rather small, peptide model, high-level theoretical descriptions can be applied to check the above structural interpretation of the observed spectral changes. For this purpose, the



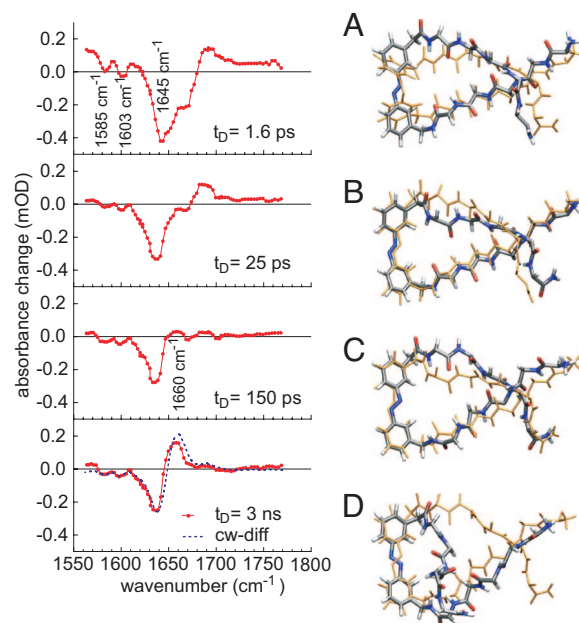
**Fig. 2.** Description of trans-cis difference spectra. (A) Probability distributions of inter- $\beta$ -strand hydrogen bonds for the cis (black dashed line) and trans (gray dashed line) equilibrium ensembles at 300 K derived from replica exchange MD simulations (see *Materials and Methods*). (B–D) Amide I bands of the model peptide calculated by DFT/MM using the INMA technique; smooth bands are obtained from line spectra by Gaussian convolution (width, 13  $\text{cm}^{-1}$ ). (B) INMA spectrum derived from three cis-hairpin snapshots. Black lines represent bands of C=O groups with inter- $\beta$ -strand hydrogen bonds; gray lines represent bands of C=O groups interacting with the solvent or with amino acid side groups or showing no hydrogen bonding. (C) INMA spectrum of three selected open trans structures. (D) Difference between the trans and cis structures. (E) Transient spectrum at  $t_0 = 25$  ps taken from Fig. 3B.

$\beta$ -hairpin unfolding in the given solvent (deuterated methanol) was first simulated by MD, and, and, subsequently, the amide I bands both before and after photoisomerization were computed by a hybrid approach combining density functional theory (DFT) with a molecular mechanics (MM) approach [for details see *Materials and Methods* and [supporting information \(SI\) Materials and Methods](#)]. Because these DFT/MM calculations require an enormous computational effort, amide I bands were calculated only for three samples of cis and trans peptide structures. These sample structures were carefully selected from large equilibrium ensembles of solvated peptide structures, which were generated by extended MD simulations of peptide-solvent systems as described in *Materials and Methods*. Our selection of cis and trans peptide sample structures was guided by the desire to learn how the amide I bands change when inter- $\beta$ -strand hydrogen bonds originally present in a nicely ordered cis- $\beta$ -hairpin structure become disrupted upon cis-trans isomerization of the azobenzene switch. This approach is supported by results of replica exchange solute tempering simulations (16) on the 300 K equilibrium conformational ensembles of the cis- and trans-hairpin, which show that the cis conformations exhibit a much larger number of inter- $\beta$ -strand hydrogen bonds than the trans states (see Fig. 24). For DFT/MM, we therefore selected from the two MD equilibrium ensembles cis-hairpin structures featuring a maximum number of interstrand hydrogen bonds and trans structures, in which all these hydrogen bonds were broken. Because the selected structures do not represent the cis and trans

equilibrium ensembles in any statistically relevant sense, the computed absolute amide I spectra do not aim at an explanation of observed stationary spectra. Instead, the difference of the computed cis and trans spectra serves to predict the changes, which are to be expected from the breaking of inter- $\beta$ -strand hydrogen bonds in a cis-trans transition.

Fig. 2*B–D* presents our DFT/MM results for the amide I bands of the selected cis (Fig. 2*B*) and trans (Fig. 2*C*) structures as well as for the trans-cis difference (Fig. 2*D*) predicting the spectral effect that is expected to result from the disruption of all inter- $\beta$ -strand hydrogen bonds. The DFT/MM spectrum of the cis-hairpins in Fig. 2*B* demonstrates that amide I bands of C=O groups involved in inter- $\beta$ -strand hydrogen bonds (black lines) are generally shifted toward frequencies that are lower than those of C=O groups interacting with the solvent or with side chains (gray lines). This classification was possible because most of the calculated amide I normal modes were found to be localized at one or two C=O groups. Because the number of strong intrapeptide hydrogen bonds is much smaller in the trans structures than in the cis structures [there are no interstrand hydrogen bonds and only two accidental intrastrand hydrogen bonds (black line)], the computed trans spectrum (Fig. 2*C*) is blue-shifted. The blue-shift is evidenced by the computed trans-cis difference spectrum of Fig. 2*D*. Thus, the shown DFT/MM results strongly support the preliminary interpretation given further above for the stationary cis-trans difference spectrum in Fig. 1 and will be important for the structural decoding of time-resolved difference spectra, to which we turn now.

**The cis-to-trans Unfolding Reaction.** For the time-resolved experiments, the unfolding process was initiated by illuminating the AMPP chromophore of the  $\beta$ -hairpin model with short light pulses (duration 700 fs) at  $\lambda = 404$  nm (17). Changes in the IR-absorption spectrum induced by the exciting light pulse were recorded by properly delayed broadband IR pulses (duration 100 fs). Because the experiment measures the time dependences in the form of difference spectra recorded as a function of time delay, we can use two representations of the data. In SI Fig. 6 we focus on the time dependence by showing a set of absorption data recorded as a function of time delay  $t_D$  at characteristic frequency positions. Here, we show in Fig. 3 a series of light-induced absorption difference spectra (red lines) recorded at various delay times after photoexcitation. In the depicted range of the amide I band, these spectra encode the differences between the transient trans peptide structures and the original cis structure. Fig. 3*A Right–D Right* present backbone snapshots that were taken from a representative MD simulation of the photoinduced unfolding at the corresponding delay times. The transient spectra in Fig. 3 show the signatures of hydrogen-bond breaking discussed above. Fig. 3*A Left* provides information for a delay time  $t_D = 1.6$  ps, when the AMPP chromophore has already acquired the new trans isomeric state. The pronounced absorption decrease peaking at  $1,645\text{ cm}^{-1}$  and the weakly enhanced absorption at  $1,690\text{ cm}^{-1}$  resemble those of the calculated spectrum (Fig. 2*D*), thus indicating disruption or weakening of hydrogen bonds and the exposure of initially hydrogen-bonded carbonyls to the solvent. Additional features of the 1.6-ps spectrum, i.e., the broadband absorption increase and decrease at the positions of the original bands of the cis ensemble at  $1,670\text{ cm}^{-1}$  (amide I) and  $1,603$  and  $1,585\text{ cm}^{-1}$  (AMPP) point to an elevated temperature of the molecule, as expected after isomerization of the excited chromophore and its internal conversion to the ground state (18). In the electronic ground state, most of the excitation energy is stored in the molecule in the form of vibrational excitation, which causes the observed absorption change. Subsequently, the excess vibrational excitation relaxes on the few-picosecond time scale. The 1.6-ps structural snapshot from a MD simulation of the light-induced hairpin unzipping (Fig. 3*A Right*) fits nicely to this interpretation. Compared with the initial cis-azobenzene- $\beta$ -hairpin (brown structure), the chromophore is in



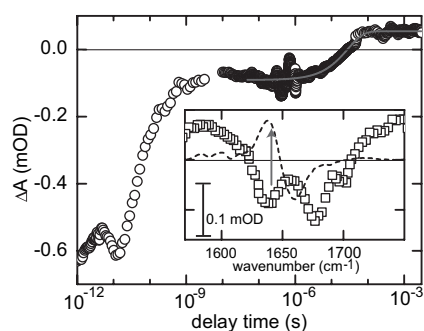
**Fig. 3.** Transient IR spectra and corresponding peptide structures. (Left) Transient absorption differences recorded at delay times of 1.6 ps (A), 25 ps (B) 150 ps (C), and 3 ns (D). The dotted curve in D is the steady-state difference spectrum calculated by subtraction of the IR spectrum of the trans ensemble from that of the cis- $\beta$ -hairpin. (Right) Snapshots (blue and gray) of the hairpin unzipping taken from a MD simulation (see SI Materials and Methods) at the same delay times as in Left. Respective precursor structures are drawn in brown. In the drawing at the top, the brown precursor is the initial cis structure.

trans conformation and the hairpin strands close to the chromophore are widened when the initial hydrogen bonds are broken. According to the simulations, a significant amount of heat remains within the peptide at this moment.

In the IR transient spectrum recorded at  $t_D = 25$  ps (see Fig. 3*B*, solid trace), there are strong indications for additional breakings of interstrand hydrogen bonds (absorption decrease shifted from  $1,645\text{ cm}^{-1}$  to  $1,637\text{ cm}^{-1}$ ) in a meanwhile thermally equilibrated peptide, which is reflected in the recovery of the broad absorption increase found at early times. The absorption increase visible around  $1,690\text{ cm}^{-1}$  indicates that a number of amide groups remains in surroundings of low polarity. Apparently, at  $t_D = 25$  ps, the molecule has reached a structure with a minimum number of interstrand hydrogen bonds. The MD simulation [see snapshots belonging to 25 ps and 1.6 ps (brown structures), respectively, in Fig. 3*B*] supports the interpretation that most of the interactions between the two strands have vanished at that time. The difference spectrum at  $t_D = 25$  ps resembles well the calculated trans-cis difference spectrum (see Fig. 2), apart from an expected overestimate of band widths inherent to the INMA approach (19, 20).

During the following period of  $\approx 100$  ps (see Fig. 3*C* and *D*), IR intensity is shifted from the  $1,690\text{-cm}^{-1}$  range toward lower frequencies, pointing to a resolution of the carbonyl groups. The transient spectrum at 150 ps exhibits solely a negative band ( $1,635\text{ cm}^{-1}$ ). At this moment, the two strands of the original  $\beta$ -sheet are largely separated, the resolution of the carbonyl groups by the solvent is complete, but the final arrangement of the trans form of the peptide is not yet reached. The incomplete relaxation is documented by the fact that the 150-ps transient absorption spectrum lacks any increased absorption near  $1,660\text{ cm}^{-1}$ , where the stationary difference spectrum (dashed lines in Figs. 3*D* and 1*B*) shows a pronounced peak. According to the MD simulations (Fig. 3*D Right*), the conformational relaxation dynamics toward the ensemble of NMR trans structures (10)



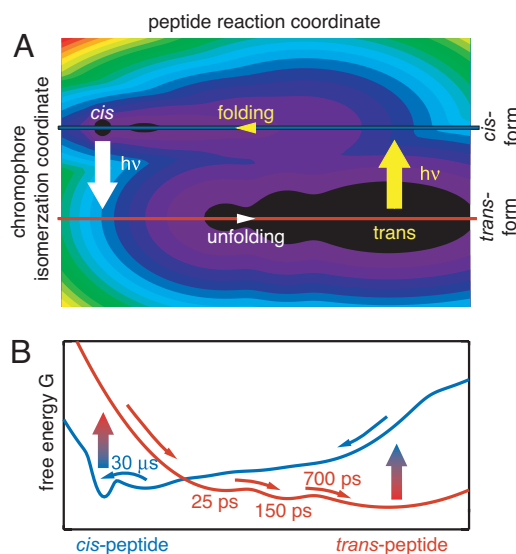


**Fig. 4.** Light-induced transient absorption changes recorded at  $1,641\text{ cm}^{-1}$  after excitation of the trans form of the model peptide at  $404\text{ nm}$  ( $t < 4\text{ ns}$ ) and  $355\text{ nm}$  ( $t > 10\text{ ns}$ ) for triggering the folding reaction. (Inset) Absorption difference spectrum recorded at a time of  $3\text{ ns}$  (symbols) and stationary trans-cis difference spectrum (dashed curve). The arrow marks the observation wavenumber of  $1,641\text{ cm}^{-1}$ .

occurs on a time scale of a few nanoseconds. Correspondingly, the positive band observed at  $1,660\text{ cm}^{-1}$  appears with a time constant of  $700\text{ ps}$  (Fig. 3B) and signifies the approach of the peptide toward the *trans*-azobenzene equilibrium ensemble. At the end of our observation time window, at  $3\text{ ns}$  (see Fig. 3D), the transient absorption difference spectrum (Fig. 3D, solid curve) is very similar to the stationary difference spectrum (Fig. 3D, dashed curve), apart from a small  $3\text{-cm}^{-1}$  frequency shift. Apparently, the peptide has nearly completed its way to the *trans*-azobenzene conformational ensemble within this very short time.

**The trans-to-cis Folding Reaction.** To analyze the reverse process, i.e., the refolding of the  $\beta$ -hairpin, transient IR spectra were recorded after illumination of the *trans*-azobenzene ensemble. A pronounced initial absorption decrease was observed around  $1,650\text{ cm}^{-1}$  (data not shown), which rapidly decays on the time scale of  $10\text{ ps}$  (see Fig. 4, data recorded at  $1,641\text{ cm}^{-1}$ ). After several nanoseconds, an absorption difference spectrum is found (compare with Fig. 4 Inset, symbols), which exhibits very small amplitudes and a shape quite different from that of the stationary trans-to-cis difference spectrum (Fig. 4 Inset, broken curve). The small amplitudes of the difference spectrum indicate that a state has been reached that resembles the initial *trans*-azobenzene peptide, i.e., has the character of a hydrophobic cluster. The subsequent absorption changes from  $10\text{ ns}$  onward have been recorded at a single probing frequency of  $1,641\text{ cm}^{-1}$ . After some artifacts due to pump-induced density fluctuations in the sample (21) on the time scale of  $10^{-7}\text{ s}$ , a  $30\text{-}\mu\text{s}$  time constant results from the transient absorption data. This transient leads from a state with reduced absorption found on the nanosecond time scale to a weak absorption increase. The data demonstrate that, in the trans-to-cis folding process, the stationary absorption is eventually reached with a time constant of  $30\text{ }\mu\text{s}$ . Thus, the intermediate state formed after photoisomerization and initial relaxation folds into a  $\beta$ -hairpin on that time scale.

It is interesting to compare the observed folding time constant of  $30\text{ }\mu\text{s}$  with results obtained with other experimental techniques. Transient temperature jump experiments in water or water guanidine hydrochloride solutions have been reported in the literature for the TrpZip2  $\beta$ -hairpin peptide, which contains the same sequence of the two strands as our model peptide (22, 23). Detection-dependent kinetics in a range below  $10\text{ }\mu\text{s}$  have been reported for TrpZip2 with fluorescence techniques that probe the hydrophobic clusters, i.e., the Trp side chains (23). Heterogeneous folding was found with stationary techniques pointing to a rough energy landscape of the unfolded peptide (24). Folding times in the  $2.5\text{-}\mu\text{s}$  range are reported from a model-based analysis of transient IR data (probing the backbone



**Fig. 5.** Free energy surfaces. (A) Two-dimensional representation of a schematic free-energy surface  $G$  plotted as a function of the isomerization coordinate of the chromophore (vertical axis) and the reaction coordinates of the cis and trans conformations of the peptide strands (horizontal axis). Light absorption (arrows) in the chromophore drives the system toward the new conformational ensemble at the bottom of the free-energy curves. (B) Horizontal cuts through the free-energy surface along the lines in A. The blue curve pertains to the cis isomeric form of the switch. The narrow minimum on the left represents the folded cis structures, and the adjacent broad minimum accounts for the unfolded cis ensemble. The red curve relates to the trans state of the switch and serves to illustrate the relaxations after the cis-to-trans photoisomerization.

carbonyls) recorded after the application of a short heating pulse (22). Apparently the temperature rise of  $\approx 10\text{--}15\text{ K}$  induced in these temperature-jump experiments causes a distortion of the folded peptide structure (25) that seems to be much weaker than the strong structural changes induced by the optical triggering process of the present  $\beta$ -hairpin model peptide. Thus the subsequent refolding process in the temperature-jump experiment requires a search only within a limited conformational space.

**Model Scheme for the Light-Triggered Reaction.** To visualize the experimentally observed light-induced structural dynamics of the  $\beta$ -hairpin peptide, we employ the schematic free-energy surface drawn in Fig. 5. In the unfolding process, light absorption and isomerization of the azobenzene transfers the cis equilibrium structures from the corresponding minima of the free-energy surface to an elevated and steep part of the free-energy surface belonging to the trans isomeric form (Fig. 5, white arrow). Here, a large mean force drives relaxations and structural changes, which lead on the time scale of  $25\text{ ps}$  to modified peptide structures with broken native hydrogen bonds. From that moment onward, the peptide structures relax toward the trans equilibrium ensemble of the azobenzene peptide. The resolution of the carbonyl groups within their new surroundings is completed within  $150\text{ ps}$ . With a time constant of  $700\text{ ps}$ , further rearrangements lead to a structural ensemble that closely resembles the *trans*-azobenzene state. In Fig. 5 the variety of structural subfamilies belonging to the trans isomer is schematically reflected by the smooth and wide minimum of the free-energy surface. In the light-triggered refolding reaction (yellow arrow), the peptide also is transferred to a steep region of the free-energy surface belonging to the cis state (Fig. 5, blue line). Fast relaxational motions guide the structural ensemble into a long-lived intermediate state characterized by a shallow local minimum. Microseconds are required to finally reach the or-

dered hydrogen-bond pattern of the cis-hairpin ensemble indicated in Fig. 5B by the narrow minimum of the blue curve.

We would like to stress the following key results of our investigation on folding and unfolding of the model peptide. (i) We have provided a structurally and temporally resolved monitoring for the complete processes of  $\beta$ -hairpin folding and unfolding. The observations reveal ultrafast dynamics with several intermediates on the subnanosecond time scale. The observed light-induced processes comprise, e.g., the subnanosecond unfolding into a hydrophobic cluster and the rearrangement of one such cluster into another that is finished on the time scale of a few nanoseconds. (ii) In our model peptide, the AMPP photoswitch is converted by light into a thermally stable  $\beta$ -turn within a few picoseconds. Corresponding reactions of local loop formation take nanoseconds in freely fluctuating peptides (26, 27). Both reactions are therefore much faster than the 30  $\mu$ s, which, according to our data, are required for the transition from a hydrophobic cluster containing the correct  $\beta$ -turn into an ordered hairpin structure. Thus, this transition appears to be the rate-limiting step of  $\beta$ -hairpin folding (compare with the discussions in refs. 28 and 29).

## Materials and Methods

**Sample Preparation.** The  $\beta$ -hairpin sample with the amino acid sequence H-Ser-Trp-Thr-Trp-Glu-AMPP-Lys-Trp-Thr-Trp-Lys-NH<sub>2</sub> was synthesized as described in ref. 10. The sample was dissolved in deuterated methanol (methanol-*d*<sub>4</sub>) (Merck, Darmstadt, Germany) to yield a concentration of 1.8 mM. The sample was held in a closed flow cycle, allowing for complete sample exchange between subsequent laser shots. The sample thickness was 220  $\mu$ m.

**Stationary IR Spectroscopy.** Stationary IR absorption and difference spectra were recorded with an IFS66 FTIR spectrometer (resolution, 2  $\text{cm}^{-1}$ ; Bruker, Ettlingen, Germany). The isomerization state of the azobenzene chromophore of the hairpin peptide was prepared by illumination of the sample via optical fibers with light at appropriate wavelengths. Switching from cis to trans was performed by light at wavelengths obtained from a KLC 2500 light source equipped with a 3-mm GG 400 filter (Schott, Mainz, Germany). For the trans-to-cis photoconversion, light from a HgXe arc lamp (LOT, Darmstadt, Germany) equipped with UG 1 and WG 320 glass filters (Schott) was used. The illumination procedure allowed us to populate the cis state to 90%. Continuous illumination of the sample with appropriate light was used to maintain the required photo-stationary state during the femtosecond time scale experiments.

**Femtosecond IR Spectroscopy.** The time-resolved experiments were performed in the visible pump, IR probe scheme. Details of the experimental set-up have been published (17). The most important features are as follows. The folding dynamics was initiated by a laser pulse centered at 404 nm with an energy of 4.4  $\mu$ J and with a pulse duration of 700 fs [spot size in the sample cell, 190  $\mu$ m (FWHM)]. The IR probe pulse was generated as previously described (17) by optical parametric amplification and difference frequency generation. The probe pulses were centered at 1,670  $\text{cm}^{-1}$ , with a spectral bandwidth of 130  $\text{cm}^{-1}$  [spot size in the sample cell, 95  $\mu$ m

(FWHM)]. The transient absorption change was recorded with a spectral resolution of 2.5  $\text{cm}^{-1}$  by a 32-channel mercury–cadmium–telluride array. The plotted absorption data represent magic angle polarization between pump and probe pulses.

**Microsecond IR Spectroscopy.** For the experiments on the microsecond time scale (see Fig. 4 *Inset*), the third harmonic of a nanosecond neodymium-doped yttrium–aluminum–garnet laser was used (355 nm) as an excitation pulse (duration, <8 ns). A single-channel IR detector recorded the transmission change of the light from a continuous IR probe laser at 1,641  $\text{cm}^{-1}$ . For details of the set-up, see ref. 30.

**Computational Methods.** The  $\beta$ -hairpin unfolding was first simulated by MD, and, subsequently, the amide I bands both before and after isomerization were computed by a hybrid approach combining DFT with a MM approach. Details of the used MD and DFT/MM methods are documented in *SI Materials and Methods*.

In brief, a periodic orthorhombic dodecahedron (inner radius, 25 Å) was filled with rigid MM methanol-*d*<sub>4</sub> models (31). The system was equilibrated by MD for 1 ns in the *NpT* ensemble using the program EGO-MMII (32). Temperature ( $T = 300$  K) and pressure ( $p = 1$  bar; 1 bar = 100 kPa) were controlled by Berendsen thermostats and barostats (33). Subsequently, the 10 best refined NMR structures (10) of the cis-AMPP hairpin were solvated by removing all overlapping methanol models followed by a 250-ps *NpT* equilibration. For the peptide, the CHARMM22 force field (34) was chosen, supplemented by DFT-derived parameters for the AMPP dye (35). Hydrogen bonds were kept rigid by M-SHAKE (36). The ultrafast ( $\approx 250$  fs), light-induced, cis–trans isomerization of the AMPP dye by MM–MD simulation was enabled by an additional model potential (2, 35). Intermediate structures obtained 100 ps after isomerization were selected as representatives for the model peptide on its way to the equilibrium ensemble of the trans conformations.

The IR bands of the cis and open trans peptide in the amide I spectral range were derived from DFT/MM calculations (37) by the protocol for an INMA given in refs. 19 and 20. For the cis case, we selected three snapshots with a maximal number of inter- $\beta$ -strand hydrogen bonds, and, for the trans case, we selected three snapshots without such hydrogen bonds from the equilibrated NMR structures. The peptide backbone building up the DFT fragment was treated with the plane-wave DFT code CPMD (38). The covalent linkages between the peptide backbone and MM fragment (dye, side chains, and solvent) were treated as described in ref. 37.

The INMA procedure yielded for each of the snapshots the vibrational frequencies and IR intensities of all backbone normal modes of the deuterated peptide. IR spectra of the backbone C=O stretching modes of the peptide were constructed by Gaussian smoothing of the obtained line spectra.

Additionally, two replica exchange solute tempering simulations (duration, 5 ns) (16) were performed for the cis and the trans isomers of the  $\beta$ -hairpin model to estimate the numbers of inter- $\beta$ -strand hydrogen bonds of the equilibrium ensembles.

We thank Birgit von Holt for supporting the microsecond experiments. The work was supported by the Deutsche Forschungsgemeinschaft (SFB 533 projects A8, B9, C1, and C3).

- Buchner J, Kiefhaber T (2005) *Protein Folding Handbook* (Wiley, Weinheim, Germany).
- Spörlein S, Carstens H, Satzger H, Renner C, Behrendt R, Moroder L, Tavan P, Zinth W, Wachtveitl J (2002) *Proc Natl Acad Sci USA* 99:7998–8002.
- Bredenbeck J, Helbing J, Sieg A, Schrader T, Zinth W, Renner C, Behrendt R, Moroder L, Wachtveitl J, Hamm P (2003) *Proc Natl Acad Sci USA* 100:6452–6457.
- Chen EF, Kumita JR, Woolley GA, Klinger DS (2003) *J Am Chem Soc* 125:12443–12449.
- Bredenbeck J, Helbing J, Kumita JR, Woolley GA, Hamm P (2005) *Proc Natl Acad Sci USA* 102:2379–2384.
- Hughes RM, Waters ML (2006) *Curr Opin Struct Biol* 16:514–524.
- Stotz CE, Topp EM (2004) *J Pharm Sci* 93:2881–2894.
- Aemissegger A, Krautler V, van Gunsteren WF, Hilvert D (2005) *J Am Chem Soc* 127:2929–2936.
- Erdelyi M, Karlen A, Gogoll A (2005) *Chem Eur J* 12:403–412.
- Dong S-L, Löweneck M, Schrader TE, Schreier WJ, Zinth W, Moroder L, Renner C (2006) *Chem Eur J* 12:1114–1120.
- Krimm S, Bandekar J (1986) *Adv Protein Chem* 38:181–364.

12. Barth A, Zscherp C (2002) *Q Rev Biophys* 35:369–430.
13. Zhuang W, Abramavicius D, Hayashi T, Mukamel S (2006) *J Phys Chem B* 110:3362–3374.
14. Kubelka J, Keiderling TA (2001) *J Phys Chem A* 105:10922–10928.
15. Torii H, Tatsumi T, Tasumi M (1998) *J Raman Spectrosc* 29:537–546.
16. Liu P, Kim B, Friesner RA, Berne BJ (2005) *Proc Natl Acad Sci USA* 102:13749–13754.
17. Schrader T, Sieg A, Koller F, Schreier W, An Q, Zinth W, Gilch P (2004) *Chem Phys Lett* 392:358–364.
18. Hamm P, Ohline SM, Zinth W (1997) *J Chem Phys* 106:519–529.
19. Schmitz M, Tavan P (2004) *J Chem Phys* 121:12233–12246.
20. Schmitz M, Tavan P (2004) *J Chem Phys* 121:12247–12258.
21. Callender RH, Dyer RB, Gilmanshin R, Woodruff WH (1998) *Annu Rev Phys Chem* 49:173–202.
22. Snow CD, Qiu LL, Du DG, Gai F, Hagen SJ, Pande VS (2004) *Proc Natl Acad Sci USA* 101:4077–4082.
23. Yang WY, Gruebele M (2004) *J Am Chem Soc* 126:7758–7759.
24. Yang WY, Pitera JW, Swope WC, Gruebele M (2004) *J Mol Biol* 336:241–251.
25. Smith AW, Chung HS, Ganim Z, Tokmakoff A (2005) *J Phys Chem B* 109:17025–17027.
26. Krieger F, Fierz B, Bieri O, Drewello M, Kiefhaber T (2003) *J Mol Biol* 332:265–274.
27. Lee JC, Gray HB, Winkler JR (2005) *J Am Chem Soc* 127:16388–16389.
28. Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) *Nature* 390:196–199.
29. Dinner AR, Lazaridis T, Karplus M (1999) *Proc Natl Acad Sci USA* 96:9068–9073.
30. Wachtveitl J, Sporlein S, Satzger H, Fonrobert B, Renner C, Behrendt R, Oesterhelt D, Moroder L, Zinth W (2004) *Biophys J* 86:2350–2362.
31. Walser R, Mark AE, van Gunsteren WF, Lauterbach M, Wipff G (2000) *J Chem Phys* 112:10450–10459.
32. Mathias G, Egwolf B, Nonella M, Tavan P (2003) *J Chem Phys* 118:10847–10860.
33. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) *J Chem Phys* 81:3684–3690.
34. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S (1998) *J Phys Chem B* 102:3586–3616.
35. Carstens H (2004) PhD dissertation (Ludwig-Maximilians-Universität, Munich).
36. Krautler V, Van Gunsteren WF, Hunenberger PH (2001) *J Comput Chem* 22:501–508.
37. Eichinger M, Grubmüller H, Heller H, Tavan P (1997) *J Comput Chem* 18:1729–1749.
38. IBM, Max Plank Institute (2001) CPMD (IBM, Armonk, NY, and Max Plank Institute, Stuttgart, Germany) Version 3.7.

# Light-triggered $\beta$ -hairpin folding and unfolding

Tobias E. Schrader\*, Wolfgang J. Schreier\*, Thorben Cordes\*, Florian O. Koller\*, Galina Babitzki\*, Robert Denschlag\*, Christian Renner†‡, Markus Loeweneck†§, Shou-Liang Dong†, Luis Moroder†, Paul Tavan\*, and Wolfgang Zinth\*¶

\*Lehrstuhl für BioMolekulare Optik and Munich Center for Integrated Protein Science, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, 80538 Munich, Germany; and †Max-Planck-Institut für Biochemie, Am Klopferspitz 18, 82152 Martinsried, Germany

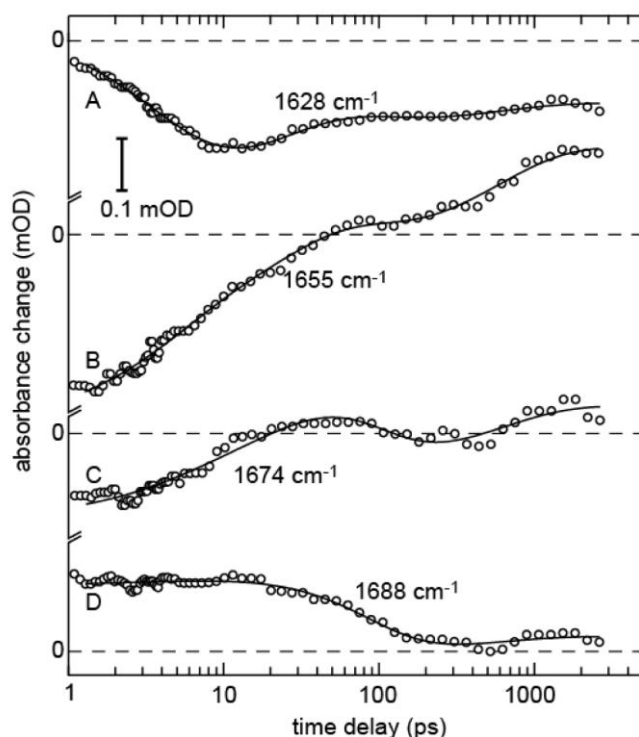
## Supporting Information

### Files in this Data Supplement:

[SI Materials and Methods](#)

[SI Figure 6](#)

[SI Figure 6](#)



**Fig. 6.** Time-dependent absorption changes (symbols) induced by the excitation pulses at 404 nm recorded as a function of the time delay at the observation wavenumbers 1,628 cm<sup>-1</sup>, 1,655 cm<sup>-1</sup>, 1,674 cm<sup>-1</sup>, and 1,688 cm<sup>-1</sup>. The solid curves were obtained by fitting the experimental data with exponential functions with the time constants 4 ps, 25 ps, 70 ps, 700 ps, and  $\infty$ .

## SI Materials and Methods

Here we describe the details of the computational methods applied to (i) simulate the photoinduced refolding of the model peptide in deuterated methanol by molecular dynamics (MD) using a molecular mechanics (MM) force field and (ii) calculate the corresponding amide I IR bands of the peptide backbone for the *cis* and *trans* isomeric states of the integrated AMPP chromophore using a hybrid force field (1), which combines density functional theory (DFT) for the peptide backbone with MM models of the amino acid side-chains and of the solvent molecules.

**Simulation System and MM-MD Methods.** A periodic orthorhombic dodecahedron (inner radius: 25 Å) was filled with 1232 rigid MM methanol-*d*<sub>4</sub> models (2). The system was equilibrated for 1 ns by MD in the *NpT* ensemble, i.e., at constant atom number *N*, pressure *p* and temperature *T*, using the program EGO-MMII (3) and controlling the temperature (*T* = 300 K) and the pressure (*p* = 1 bar) by Berendsen thermostats and barostats (4), respectively [coupling constants, 0.5 ps (*T*) and 5 ps (*p*); isothermal compressibility,  $12 \cdot 10^{-5} \text{ bar}^{-1}$  (2)]. Here, the equations of motion were integrated by a multiple time-step algorithm (5) with a basic time step of 1 fs. The long-range electrostatics was treated by fast hierarchical multipole expansions (6) combined with a moving boundary reaction field approach (3), where a dielectric continuum [ $\epsilon = 32.6$  (2)] was assumed at distances beyond 25 Å. Lennard-Jones interactions were cut off at 10 Å and a long-range correction to the Lennard-Jones energy was applied (7).

Subsequently, the 10 best refined NMR structures (8) of the *cis*-AMPP hairpin were solvated by removing all overlapping methanol models. For the peptide the CHARMM22 force field (9) was chosen and supplemented by DFT derived parameters for the AMPP dye and for its linkage to the peptide (10). The peptide-solvent systems were carefully equilibrated by 250-ps *NpT* simulations keeping all covalent bonds involving hydrogen atoms rigid by M-SHAKE (relative tolerance,  $10^{-6}$ ) (11). The resulting ensemble of solvated *cis*-hairpin structures served as a starting point for MM-MD simulations of the light-induced *cis*-*trans* isomerization and for computing the IR spectra of the *cis*-hairpin in methanol by DFT/MM.

In addition, the 300 K *cis*- and *trans*-hairpin equilibrium ensembles were estimated by two REST MD simulations (12) of 5 ns length, in which eight replicas covered the temperature range from 300 K up to 500 K (300 K, 322 K, 346 K, 372 K, 401 K, 432 K, 465 K, 500 K). Here, the starting structures were the eight best refined NMR structures which were additionally equilibrated for 100 ps to adopt the designated temperatures. After equilibration, exchanges between neighboring replicas were attempted every 2 ps as described in Ref. (13) with an average acceptance ratio of  $\approx 20\%$ . Inter- $\beta$ -strand H-bond statistics were derived from the 300 K results.

**MM-MD Simulations of the Light-Induced *cis*-*trans* Relaxation.** The MM-MD simulations of the light-induced *cis*-*trans* photoisomerization of the AMPP chromophore and of the subsequent peptide relaxation from the *cis* into the *trans* conformational equilibrium ensemble were enabled by a model potential (10, 14) driving an inversion reaction at one of the atoms within the central N=N bond of AMPP. That model potential drives an ultrafast ( $\approx 250$  fs) *cis*-*trans* isomerization of AMPP and deposits an energy into the molecule, which equals that of the absorbed photon (70.7 kcal/mol). Each of the equilibrated *cis*-hairpin structures was simulated for 1 ps in the NVE ensemble using now the Verlet algorithm (15) for integration until the photoisomerization was triggered by switching on the quoted model



potential. The NVE ensemble was maintained for the following 10 ps. Then the solvent molecules were coupled to a Berendsen thermostat ( $T = 300$  K) and NVT simulations were continued for another 3,000 ps applying the multiple time-step integrator (5). The coordinates, kinetic and potential energies of the peptide were recorded every 5 fs during the first 10 ps, every 50 fs during the next 90 ps, and every 500 fs during the remainder of the simulation for analysis of energy dissipation and structural relaxation dynamics. Intermediate structures obtained 100 ps after photoisomerization were selected as representatives for the model peptide on its way to the equilibrium ensemble of trans-conformations.

**DFT/MM Calculations of Amide I Bands.** The IR bands of the *cis* and *trans* peptide in the amide I spectral range were derived from DFT/MM calculations (1) by the protocol for an "instantaneous normal mode analysis" (INMA) described in refs. 16 and 17. INMA requires snapshots of the peptide solvent system. For the *cis*-case we selected three snapshots with a maximal number of inter- $\beta$ -strand H-bonds and for the *trans*-case three snapshots without such H-bonds from ensembles of equilibrated NMR structures. For the INMA calculations the solvent-solute snapshots were partitioned into DFT and MM fragments by assigning in the *cis*-case the complete peptide backbone and in the *trans*-case the backbone of one of the two isolated branches (residues 1-5 or 6-10, treated independently) to the DFT fragment, resulting in 78 (*cis*) and  $\approx 40$  (*trans*) atoms to be treated by DFT. To enable the application of the plane-wave DFT code CPMD (18) the DFT fragments were enclosed by rectangular boxes such that each DFT atom was located at least 3 Å away from the next box surface (e.g., resulting in a size of  $24 \times 14 \times 14$  Å<sup>3</sup> for *cis*). A plane wave cut-off of 70 Ry, the gradient-corrected exchange functional of Becke (19), the correlation functional of Perdew (20), and the norm-conserving pseudopotentials of Troullier and Martins (21) were applied. The covalent linkages between the peptide backbone and MM fragment were treated as described in ref. 1.

In the INMA computation of the amide I bands the solvent cages of the structural snapshots were kept rigid and the DFT/MM energies of the systems were minimized by relaxing the coordinates of the azopeptide until the maximal gradient fell below 1.0 kcal/(mol·Å). At the optimized geometries the Hessians were calculated for all DFT atoms by symmetric finite differences ( $\Delta x = 0.01$  Å) of analytical DFT/MM forces. The amide groups of the peptide backbone were chosen to be deuterated as in the experimental setting. The INMA procedure yielded then for each of the snapshots the vibrational frequencies and IR intensities of all backbone normal modes. IR spectra of the backbone C=O stretching modes in the *cis* and *trans* conformations of the peptide were constructed by superposing the obtained line spectra and convolving them with a Gaussian of 13 cm<sup>-1</sup> width.

**Time-Dependent Absorption Changes at Distinct Probing Wavenumbers.** The time dependence of the absorption changes allows to obtain insight into the time scales of specific processes that occur during the unfolding reaction (see SI Fig. 6). It reflects transients on different time scales: Directly after the photoisomerization of the AMPP optical switch, which is finished within a few picoseconds, the absorption is changed over the complete recorded spectral range. Subsequent strong transient absorption changes occur with a time constant of  $\approx 4$  ps. They lead to a reduced absorption on the far low frequency side of the amide I band (see curve A, 1,628 cm<sup>-1</sup>). The strong initial absorption decrease observed at higher frequencies recovers partially with the same time constant (trace B). Subsequently there is a weak transient on the 20 ps and 70 ps time scale (traces A, C, and D). A pronounced absorption increase is found at 1,655 cm<sup>-1</sup> (trace C) which occurs with a time constant of  $\approx 700$  ps. The absorption changes reached at the end of the observation period fit

well to those of the stationary difference spectrum (Fig. 1B, dashed curve). A fit performed with the presented set of time constants describes well the recorded absorption data.

1. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) *J Chem Phys* 110:10452-10467.
2. Walser R, Mark AE, van Gunsteren WF, Lauterbach M, Wipff G (2000) *J Chem Phys* 112:10450-10459.
3. Mathias G, Egwolf B, Nonella M, Tavan P (2003) *J Chem Phys* 118:10847-10860.
4. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) *J Chem Phys* 81:3684-3690.
5. Eichinger M, Grubmüller H, Heller H, Tavan P (1997) *J Comput Chem* 18:1729-1749.
6. Niedermeier C, Tavan P (1994) *J Chem Phys* 101:734-748.
7. Allen MP, Tildesley D (1987) *Computer simulations of liquids* (Clarendon, Oxford).
8. Dong S-L, Löweneck M, Schrader TE, Schreier WJ, Zinth W, Moroder L, Renner C (2006) *Chem Eur J* 12:1114-1120.
9. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S (1998) *J Phys Chem B* 102:3586-3616.
10. Carstens H (2004) *Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modellbildung* (Ludwig-Maximilians-Universität, Munich).
11. Krautler V, Van Gunsteren WF, Hunenberger PH (2001) *J Comput Chem* 22:501-508.
12. Liu P, Kim B, Friesner RA, Berne BJ (2005) *Proc Natl Acad Sci USA* 102:13749-13754.
13. Sugita Y, Okamoto Y (1999) *Chem Phys Lett* 314:141-151.
14. Spörlein S, Carstens H, Satzger H, Renner C, Behrendt R, Moroder L, Tavan P, Zinth W, Wachtveitl J (2002) *Proc Natl Acad Sci USA* 99:7998-8002.
15. Verlet L (1967) *Physical Review* 159:98-103.
16. Schmitz M, Tavan P (2004) *J Chem Phys* 121:12233-12246.
17. Schmitz M, Tavan P (2004) *J Chem Phys* 121:12247-12258.
18. IBM, Max Plank Institute (2001) CPMD (IBM, Armonk, New York, and Max Plank Institute, Stuttgart, Germany) Version 3.7.
19. Becke AD (1988) *Phys Rev A* 38:3098-3100.
20. Perdew JP (1986) *Phys Rev B* 33:8822-8824.

21. Troullier N, Martins JL (1991) *Phys Rev B* 43:1993-2006.



## 4 Zusammenfassung und Ausblick

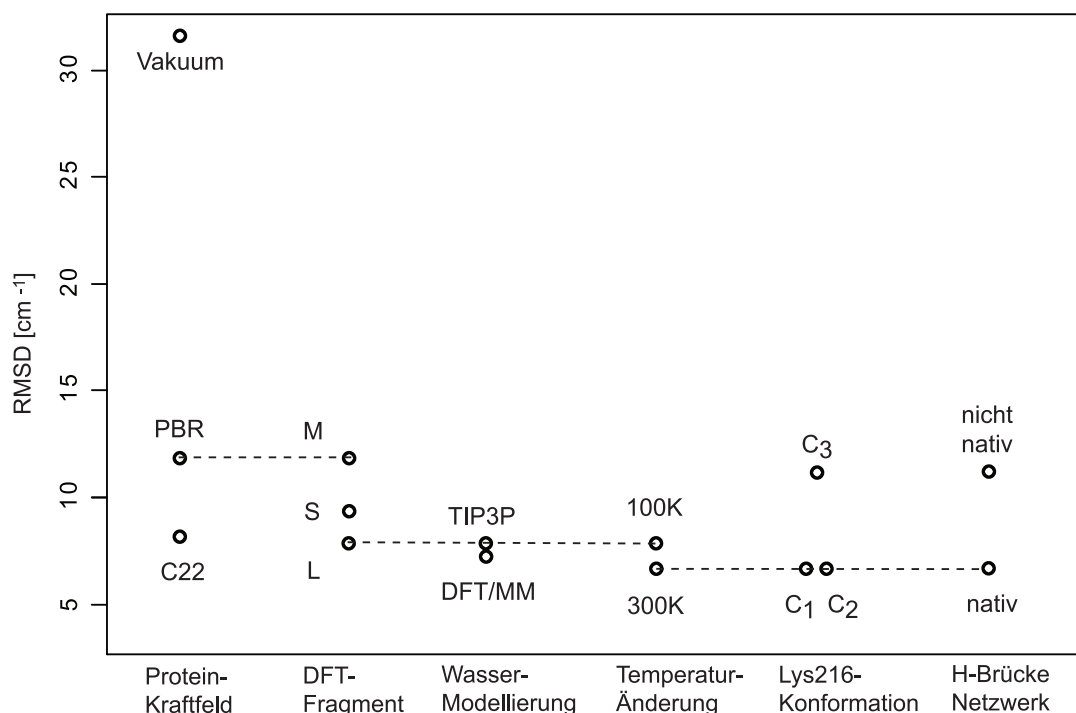
Wie in der Einleitung bereits ausführlich beschrieben wurde, steuert die Elektrostatik der Proteinumgebung die Eigenschaften des Retinalchromophores von BR. Das große und stark polarisierbare  $\pi$ -Elektronensystem der RSBH<sup>+</sup> reagiert auf kleinste Veränderungen der Umgebungselektrostatik. Die Berechnung ihrer Schwingungsspektren stellt somit die folgenden Anforderungen an die verwendete theoretische Methode: wegen der hohen Polarisierbarkeit der RSBH<sup>+</sup> muss sie sehr präzise sein und wegen ihrer Größe muss sie effizient sein. Die DFT/MM Instantane Normalmodenanalyse (INMA) entspricht diesen Anforderungen. Sie ist am experimentellen Szenario orientiert, indem sie mittels einer MM-MD Simulation ein Ensemble von Strukturen generiert und für jede Struktur des Ensembles das Anregungsspektrum berechnet [91, 92].

Es ist jedoch seit langem bekannt und hat sich in unseren Simulationen erneut gezeigt, dass die Kristallstruktur der BR-Bindungstasche in MM-MD Simulationen mit den bei Proteinsimulationen üblichen nicht-polarisierbaren MM-Kraftfeldern (CHARMM22, AMBER, GROMACS) nicht stabil bleibt [93, 99, 104]. Die Wasserstoffbrücke zwischen der RSBH<sup>+</sup> und dem Wassermolekül 402, wird gebrochen. Aufgrund der sehr viel längeren Dauer unserer CHARMM22 Simulationen wurde dort eine weitere Instabilität offenbar: Gelegentlich änderte sich die Konformation der aliphatischen Kette von Lys216. Um zu überprüfen, ob die strukturellen Instabilitäten verschwinden, wenn die elektronische Polarisierung der Chromophorbindungstasche im Mittel berücksichtigt wird, habe ich die Partialladungen derjenigen Seitengruppen, die sich in der Nähe des Chromophors befinden, mit einer DFT/MM Methode neu ermittelt und damit das polarisierte Kraftfeld PBR erzeugt. Bei 50 ns überdeckenden MD Simulationen mit dem PBR-Kraftfeld, blieben die Geometrien des Wasserstoff-Brückennetzwerks in der Chromophorbindungstasche und der Lysinkette dicht an ihren kristallographisch ermittelten Struktur. Dieser Befund bestätigt die z.B. von Ponder et al. [105] und Tavan et al. [97] formulierte Vermutung, dass die Vernachlässigung von Polarisierungseffekten in herkömmlichen molekülmechanischen Kraftfeldern zu schwerwiegenden Artefakten in der Struktur und der Dynamik von Proteinen führen kann.

Sollten neben der bei tiefen Temperaturen beobachteten kristallographischen Konformation noch andere Konformationen der Lysinkette im 300 K Ensemble existieren, so sind diese, wie es aus den 50 ns Simulationen folgt, durch hohe Freie-Energie Barrieren von der kristallographischen Konformation getrennt. In diesem Fall kann man nicht erwarten, dass in der Zeit, die man mit konventionellen Simulationsmethoden abdecken kann, solche Konformationsübergänge beobachten werden können, so dass hier auf alternative Techniken zum Abtasten des Konformationsraums, wie die Hamilton Replica Exchange (HRE) Methode [106] zurückgegriffen werden musste. Die durchgeführten HRE-Simulationen haben gezeigt, dass die Lysinkette bei Raumtemperatur drei Hauptkonformere aufweist  $C_i$ ,  $i \in \{1, 2, 3\}$ , die unterschiedliche Temperatureigenschaften besitzen. Die kristallographische Konformation  $C_1$  dominiert bei 300 K mit der Wahrscheinlichkeit von 58 % und ist enthalpisch begünstigt. Ihre Wahrscheinlichkeit fällt dementsprechend mit steigender Temperatur ab. Im Gegenteil dazu ist die  $C_2$  Konformation entropisch begünstigt. Sie hat eine Wahrscheinlichkeit von 32 % bei 300 K und wird erst bei höheren Temperaturen dominant. Die am schwächsten besetzte Konformation  $C_3$  (10 %) ist dagegen ganz von der Temperatur unabhängig.

Wie ich in der Einleitung erwähnt habe, zeigte die NMR-Strukturanalyse [74] Flexibilität in der Lysinkette. Die von mir durchgeführte hierarchische Klassifikation der NMR Strukturen hat nun gezeigt, dass neun der zwölf Strukturen als  $C_2$  klassifiziert werden und die restliche drei in der  $C_1$  Konformation vorliegen. Vor dem Hintergrund der in der Einleitung angesprochenen Prozeduren zur NMR-Strukturbestimmung ist die Dominanz der entropisch begünstigten Strukturen im NMR-Ensemble nicht überraschend, und zeigt lediglich, dass die NMR-Strukturen Spuren des angewendeten Optimierungsverfahrens aufweisen. Somit klären die durchgeführten HRE-Simulationen die bestehenden Diskrepanzen zwischen NMR- und Kristallstrukturen, und bestätigen außerdem die bereits von Diller und Stockburger [20] anhand kinetischer Untersuchungen gemachte Aussage, dass BR bei 300 K heterogen ist.

Nach der Erstellung stabiler MD-Trajektorien zu jeder Konformation  $C_i$  habe ich mittels der von Eichinger et al. [56] implementierten DFT/MM Methode die Schwingungsspektren des Chromophors in verschiedenen Schnappschüssen berechnet. Um die möglichen Effekte der Proteinumgebung und dem verwendeten DFT-Fragment zu untersuchen, habe ich zusätzlich die Schwingungsspektren des Retinals im Vakuum und in der 100 K Kristallstruktur berechnet, für die ich unterschiedliche Kraftfelder und unterschiedliche Aufteilungen in DFT- und MM-Fragment verwendet habe. Durch anschließende ausführliche Analyse der durchgeführten Berechnungen konnte somit eine Reihe wichtiger methodischer Fragestellungen geklärt werden, die ich nun kurz anhand der Abbildung 4.1 diskutieren will.



**Abbildung 4.1:** Abweichungen der berechneten IR-Spektren vom Experiment bei unterschiedlichen Modellierungen des BR.

Zum Beispiel weiß man aus [107], dass sich die IR- Spektren von mittelgroßen Molekülen im Vakuum mit sehr hoher Genauigkeit und vergleichbar geringerem Aufwand berechnen lassen. So kann man sich die Frage stellen: Kann man das Schwingungsspektrum des Retinals anhand der Vakuumrechnungen aufklären?

Die berechneten Streckschwingungen des Chromophors im Vakuum zeigen eine deutliche Rotverschiebung der Linien gegenüber der experimentellen Beobachtungen. Diese Rotverschiebung ist auf die stark delokalisierte Ladungsverteilung des Retinals im Vakuum und damit auch die planare Struktur zurückzuführen. Wie ich bereits erwähnt habe, wird die Struktur und die Ladungsverteilung des Retinals entscheidend durch die Einbettung ins Protein verändert. Im Einklang damit zeigen auch die Schwingungsbanden des Retinals in der Proteinmatrix aus der Kristallstrukturanalyse bei der Modellierung mit PBR- und C22-Kraftfeldern, erheblich kleinere Abweichungen von Experiment.

Somit ist klar, dass es für die Dekodierung des IR-Spektrums des Retinals nicht ausreicht, sein Schwingungsspektrum im Vakuum zu analysieren. Stattdessen sollten mit einer geeigneten Aufteilung des Systems in das DFT- und MM-Fragment Hy-

bridrechnungen in der Proteinumgebung durchgeführt werden. Doch zunächst soll geklärt werden, wie groß das DFT-Fragment gewählt werden soll und ob die Wahl des DFT-Fragments einen Einfluss auf die Qualität der berechneten Schwingungsbanden hat?

Die Variierung der Größe des DFT-Fragmentes in der Kristallstruktur hat gezeigt, dass die größte Abweichung bei der Streckschwingungen der Doppelbindungen im Retinal erfolgt bei dem mittelgroßen der gewählten DFT-Fragmente (Modell M: Ret-CH<sub>2</sub>-CH<sub>3</sub>). Für das kleinstmögliche DFT-Fragment S, das nur das Retinal und C<sub>ε</sub>-Gruppe der Lysinkette enthält wird die Abweichung kleiner, und für das DFT-Fragment L, in das die Lys216-Seitenkette und Lys216-Peptidplättchen eingeschlossen wurden, wird sie minimal. Somit liefert das größte DFT-Fragment offensichtlich eine bessere Beschreibung für das Retinal im BR.

Außerdem spielt für die Modellierung des Retinals im BR außer dem Kraftfeld für das Protein auch das verwendete Wassermodell eine sehr wichtige Rolle. Aus diesem Grund sollte als weiteres geklärt werden, ob das verwendete TIP3P Wassermodell hinreichend gut für die Beschreibung der internen Wassermoleküle im BR ist.

Trotz der ersten Erwartung, dass das Dipolmoment bei dem TIP3P Wassermodell die obere Grenze für die Polarität des Wassers in der Proteinumgebung darstellt, haben DFT/MM-Rechnungen in der Kristallstruktur gezeigt, dass das Dipolmoment bei jedem der drei internen Wassermoleküle diesen Wert um etwa 10% überschreitet und das Schwingungsspektrum des Retinals in der Kristallstruktur bei der Verwendung des mehr polaren Wassermodells noch besser das experimentelle Spektrum reproduziert (Modell DFT/MM in der Abb./4.1). Diese Beobachtungen legen den Schluss nahe, dass die Verwendung polarisierbaren Wassermodellen wesentliche Vorteile für die Modellierung des BR verspricht: erstens, aufgrund des größeren Dipolmoments wird die Kristallstruktur in den MM/MD-Simulationen noch stärker stabilisiert; zweitens, die Abweichungen zwischen experimentellen Daten und INMA Spektren werden wahrscheinlich noch kleiner. Da diese Ergebnisse aber erst nach der Erstellung der MD-Trajektorien gewonnen wurden, wurde das mehr polare Wassermodell im Weiteren nicht verwendet.

Wie ich bereits erwähnt habe, wurden alle oben diskutierte Berechnungen der Schwingungsspektren nur für die Kristallstruktur bei 100 K durchgeführt. An dieser Stelle kann nun die nächste interessante Frage gestellt werden: Ändern sich die Bandenlagen, wenn man die Schwingungsanalyse für das Ensemble der Raumtemperaturstrukturen durchführt?



---

Da aus der oben beschriebenen Berechnungen klar geworden war, dass man die beste Beschreibung für das Retinal erreicht, wenn man das größte DFT-Fragment mit dem PBR-Kraftfeld kombiniert, habe ich die INMA-Rechnungen ausschließlich für diese Modellierung durchgeführt. Die Schwingungsanalyse hat gezeigt, dass für das thermische Ensemble der BR-Strukturen bei 300 K die mittlere Abweichung von experimentellen Werten kleiner ist als für die Kristallstruktur bei 100 K, was darauf hindeutet, dass die Berücksichtigung der thermischen Fluktuationen in der BR-Struktur durch MD-Simulation eine realistischere Modellierung des Retinals vermittelt.

Wie man jedoch aus der vorhergehenden HRE-Simulation weiß, besteht das thermische Ensemble der BR-Strukturen bei 300 K nicht nur aus der kristallographischen Konformation der Lysinkette. Demzufolge gehört zu einer realistischen Beschreibung des Retinals im BR auch die Berücksichtigung der zwei anderen Konformationen der Lysinkette. Somit stellte sich hier die Frage, in wie weit die zwei weiteren Lysinkonformationen die Schwingungsspektren des Retinals beeinflussen?

Es war ein interessantes und unerwartetes Ergebnis, dass trotz der unterschiedlichen Temperatureigenschaften die am stärksten besetzten Konformationen ( $C_1$  und  $C_2$ ) sehr ähnliche Schwingungsspektren haben. Die Konformation  $C_3$ , die im thermischen Ensemble sehr schwach besetzt ist, zeigt dagegen eine starke Rotverschiebung der Schwingungsbanden und unterscheidet sich damit deutlich von den beiden dominierenden Konformationen. Dieser Befund wirft nun die Frage auf, welche Strukturveränderungen im BR man anhand der Schwingungsanalyse des Retinals unterscheiden kann?

Die  $C_1$  und  $C_2$  Konformationen der Lysinkette kann man offensichtlich anhand der Schwingungsanalyse des Retinals nicht unterscheiden. Zwar wird die Struktur des Retinals sowie seine Ladungsverteilung signifikant durch den Übergang von  $C_1$  zu  $C_2$  beeinflusst, jedoch sieht das Schwingungsspektrum des Retinals in beiden Konformationen recht ähnlich aus. Die Konformation  $C_3$  der Lysinkette hat zwar eine ähnliche Auswirkung auf die Struktur des Retinals wie die Konformation  $C_2$  der Lysinkette, verändert dagegen das Schwingungsspektrum des Chromophores stark. Man sieht damit, dass nicht alle Konformationen der Lys216-Kette das Schwingungsspektrum des Retinalchromophores verändern.

Als nächstes konnte überprüft werden, ob die Fehlorientierung der Schiffchen Base des Retinals die in der Simulationen mit CHARMM-Kraftfeld entsteht, sich in den Spektren des Retinals sichtbar macht und ob man aufgrund der Schwingungsspektren feststellen kann, ob diese Konformation doch eine mögliche Struktur des Retinals darstellt. Die berechneten IR Spektren zeigen, dass die durch das CHARMM-Kraftfeld verursachte fehlerhafte Orientierung der Schiff-Base (nicht-nativ-Modell in der

Abb. 4.1) zu einer Rotverschiebung seiner Spektren führt. Das betrifft insbesondere die C=N-Mode des Retinals, die im experimentellen IR-Spektrum sehr klar bei  $1642\text{ cm}^{-1}$  zu sehen ist. Bei den Strukturen aus dem CHARMM-Ensemble taucht sie jedoch bei  $1632\text{ cm}^{-1}$  auf, was sich in der größeren Abweichung vom Experiment als bei der nativen Orientierung der Schiff-Base widerspiegelt. In diesem Fall kann man also eine deutliche spektrale Verschiebung sehen und anhand der Schwingungsanalyse eindeutig feststellen, dass es sich dabei um einen Artefakt des verwendeten Kraftfeldes handelt.

Wie ich bereits in der Einleitung erklärt habe, soll die vorliegende Arbeit der Beantwortung der Frage beitragen, mit welcher Qualität die IR-Spektren des Retinals bei seiner sorgfältigsten Modellierung berechnet werden können?

Durch systematische Analyse der möglichen Einflussparameter konnte ich feststellen, dass man die beste Modellierung des Retinals erreicht, wenn man Polarisierungseffekte der Proteinumgebung berücksichtigt und das DFT-Fragment verwendet, das das Retinal und die Lys216 umschließt. Berücksichtigt man die zwei bei 300 K dominierende Konformationen der Lysinkette, so beträgt die Abweichung der berechneten Streckschwingungen der Doppelbindungen von den experimentellen Werten lediglich  $6.5\text{ cm}^{-1}$ .

Ein weiteres Ziel meiner Arbeit war, Daten aus den zeitaufgelösten Ultrakurzzeit-Spektroskopie-Experimenten am  $\beta$ -Hairpinpeptid mit theoretischen Daten zu vergleichen und die Entfaltungsdynamik des  $\beta$ -Hairpinpeptides nach der Anregung des Photoschalters im cis-Zustand aufzuklären.

Der Entfaltungsprozess konnte zwar mit MD simuliert werden, jedoch konnte nachfolgend wegen extrem langen Rechenzeiten die AmidI-Banden lediglich für drei Strukturen vor der Isomerisierung des Photoschalters und drei Strukturen nach seiner Isomerisierung berechnet werden.

Trotzdem konnte durch die gute Übereinstimmung der berechneten IR Spektren mit den experimentellen Spektren ein Faltungsmodell für die  $\beta$ -Hairpin Struktur aufgestellt werden. Die berechnete DFT/MM Spektren zeigen deutlich, dass die AmidI-Banden der Gruppen, die mit dem benachbarten Peptidstrang Wasserstoffbrücken ausbilden, bei kleineren Frequenzen liegen als die AmidI-Banden der Gruppen, die mit dem Lösungsmittel kontaktieren. Diese spektrale Verschiebung reproduziert sehr genau den Effekt den man mit den Methoden ultraschnellen Spektroskopie beobachtet und bestätigt damit auch die Aussagen von [88, 91, 92, 99], dass die DFT/MM-Methode den Lösungsmiteleinfluss sehr genau erfassen kann.

---

Andererseits zeigen die durchgeführten Rechnungen die Beschränkungen des DFT/MM-Ansatzes: die Berechnung der AmidI-Banden mit DFT/MM-Methode ist sogar bei so einem kleinen Modellsystem wie  $\beta$ -Hairpinpeptid enorm aufwendig und die Anwendung der Methode auf größere Peptide ist damit nicht möglich. Zwar wurde auf dem Gebiet der Entwicklung von effizienten und gleichzeitig genauen Algorithmen zur Berechnung der Amid-Banden eine große methodische Arbeit geleistet [108], bedürfen jedoch einige methodische Konzepte immer noch einer Weiterentwicklung, so dass die Anwendung der DFT/MM-Methode derzeit unumgänglich ist.

Insgesamt liefert die vorliegende Arbeit eine Vielzahl von interessanten Ergebnissen, die man anhand der MD-Simulationen kombiniert mit DFT/MM-Rechnungen machen konnte: zum einen konnte man eine Reihe methodischer Fragen bezüglich der Berechnung der Schwingungsspektren des Retinalchromophores im Bakteriorhodopsin beantworten und Diskrepanzen zwischen den Aussagen von NMR- und Röntgenstrukturanalyse aufklären; zum anderen konnte man die Enfaltungsdynamik des  $\beta$ -Hairpinpeptids aufklären und die Limitierungen in den Anwendungsmöglichkeiten der DFT/MM-Methode zeigen.

Ausgehend von den Ergebnissen dieser Arbeit kann nun versucht werden, die Schwingungsspektren weiterer Intermediaten im Photozyklus zu berechnen. Man kann jetzt davon ausgehen, dass insbesondere die Schwingungsanalyse des L-Intermediats viele weitere Einsichten in den Protonentransportmechanismus vermitteln kann. Da aber die Kristallstruktur des L-Intermediats noch nicht ausreichend aufgeklärt ist, muss man hier MD-Simulationen unter Annahme unterschiedlichen Photoisomerisierungsmechanismen durchführen. Diese sollten auf der Zeitskala von einigen Mikrosekunden erfolgen, um die Umorientierung der Wassermoleküle in der Umgebung des Retinals zu erlauben. Außerdem erfordert die Berechnung der Schwingungsspektren des Retinals eine präzise Beschreibung der Umgebungselektrostatik, die hinreichend genau erst mit der Entwicklung polarisierbaren Kraftfeldern erfolgen kann. Ferner verspricht auch den Einsatz der schnelleren DFT-Programmen wie zum Beispiel das Programm Quickstep [107] einen wesentlich geringeren Rechenaufwand für die Schwingungsanalyse *in situ*.

Außerdem könnten die vorliegenden Rechnungen am  $\beta$ -Hairpinpeptid zur Weiterentwicklung der effizienten Methoden zur Berechnung der Amid-Banden bei Peptiden und Proteinen beitragen. Dabei könnten viele Informationen über die Positionen der Schwingungsbanden im  $\beta$ -Hairpinpeptid aus DFT/MM-Rechnungen zum fundierten Vergleich mit den Ergebnissen der jeweiligen Methode verwendet werden. Damit könnten neue Konzepte zur Dekodierung der Amid-Banden entwickelt werden, dessen Anwendung enorme Erfolge in der Strukturaufklärung verspricht.



# Literaturverzeichnis

- [1] Babitzki, G., R. Denschlag und P. Tavan. Polarization effects stabilize bacteriorhodopsin's chromophore binding pocket: A molecular dynamics study. *J. Phys. Chem. B* **113**, 10483–10495 (2009).
- [2] Babitzki, G., G. Mathias und P. Tavan. The infra-red spectra of the retinal chromophore in bacteriorhodopsin calculated by a DFT/MM approach. *J. Phys. Chem. B* **113**, 10496–10508 (2009).
- [3] Schrader, T. E., W. J. Schreier, T. Cordes, F. O. Koller, G. Babitzki *et al.* Light triggered  $\beta$ -hairpin folding and unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15729–15734 (2007).
- [4] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- [5] Watson, J. D. und F. H. C. Crick. A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
- [6] Branden, C. und J. Tooze. *Introduction to Protein Structure*. Garland Publishing, Inc., New York (1991).
- [7] Dawson, P., T. Muir, I. Clark-Lewis und S. Kent. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
- [8] Zhang, C. und S. Kim. Overview of structural genomics: from structure to function. *Current opinion in chemical biology* **7**, 28–32 (2003).
- [9] Chan, H. S. Kinetics of protein folding (note). *Nature* **373**, 664–665 (1995).
- [10] McCammon, J. A. und S. C. Harvey. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge (1987).
- [11] Dobson, C. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
- [12] Oesterhelt, D. und W. Stoeckenius. Functions of a New Photoreceptor Membrane. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2853–2857 (1973).
- [13] Grigorieff, N., T. A. Ceska, K. H. Downing, J. M. Baldwin und R. Henderson. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).
- [14] Essen, L.-O., R. Siegert, W. D. Lehmann und D. Oesterhelt. Lipid patches in membrane protein oligomers. Structure of the bacteriorhodopsin-lipid complex. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11673–11678 (1998).
- [15] Luecke, H., B. Schobert, H.-T. Richter, J.-P. Cartailler und J. K. Lanyi. Structure of Bacteriorhodopsin at 1.55 Å Resolution. *J. Mol. Biol.* **291**, 899–911 (1999).

- [16] Tavan, P., K. Schulten, W. Gärtner und D. Oesterhelt. Substituents at the C<sub>13</sub> position of retinal and their influence on the function of bacteriorhodopsin. *Biophys. J.* **47**, 349–355 (1985).
- [17] Tavan, P. und K. Schulten. Evidence for a 13,14-cis cycle in bacteriorhodopsin. *Biophys. J.* **81**, 81–89 (1986).
- [18] Großjean, M. F. und P. Tavan. Observation of ‘N’ in the photocycle of bacteriorhodopsin: A satisfactory confirmation of theoretical predictions. In A. Pullman (Herausgeber), *Transport through Membranes: Carriers, Channels and Pumps*, Seiten 399–408. D. Reidel Publishing Company, Dordrecht: Holland (1988).
- [19] Fahmy, K., M. F. Großjean, F. Siebert und P. Tavan. The photoisomerization in bacteriorhodopsin studied by FTIR linear dichroism and photoselection experiments combined with quantumchemical theoretical analysis. *J. Mol. Spectrosc.* **214**, 257–288 (1989).
- [20] Diller, R. und M. Stockburger. Kinetic Resonance Raman Studies reveal different Conformational States of Bacteriorhodopsin. *Biochemistry* **27**, 7641–7651 (1988).
- [21] Zinth, W. und D. Oesterhelt. The Primary Photochemical Process in Bacteriorhodopsin. *Photobiology* Seiten 531–535 (1991).
- [22] Osterhelt, M. und W. Stockenius. Functions of a new photorezeptor membrane. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2853–2857 (1973).
- [23] <http://www.chemie.uni-marburg.de>.
- [24] Sumper, M., H. Reitmeier und D. Oesterhelt. Zur Biosynthese der Purpurmembran von Halobakterien. *Angew. Chem* **7**, 203–210 (1976).
- [25] Dunn, R., N. Hackett, K. Huang, S. Jones, H. Khorana *et al.* Studies on the light-transducing pigment bacteriorhodopsin. In *Cold Spring Harb Symp Quant Biol*, Band 48, Seiten 853–862 (1983).
- [26] Schmidt, R. F. und G. Thews. *Physiologie des Menschen*. Springer-Verlag, Berlin (1985).
- [27] Mathies, R. A., S. W. Lin, J. B. Ames und W. T. Pollard. From Femtoseconds to Biology: Mechanism of Bacteriorhodopsin’s Light-Driven Proton Pump. *Ann. Rev. Biophys. Biochem.* **20**, 491–518 (1991).
- [28] Siebert, F., W. Mäntele und W. Kreutz. Evidence for the protonation of two internal Carboxylic Groups during the Photocycle of Bacteriorhodopsin. *FEBS Letters* **141**, 82–87 (1982).
- [29] Gerwert, K., B. Hess, J. Soppa und D. Oesterhelt. Role of Aspartate-96 in Proton Translocation by Bacteriorhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4943–4947 (1989).
- [30] Coleman, M. *et al.* Asp 46 can substitute for Asp 96 as the Schiff base proton donor in bacteriorhodopsin. *Biochemistry* **34**, 15599–15606 (1995).
- [31] Dioumaev, A. K. Infrared Methods for the Protonation State of Carboxylic Amino Acids in the Photocycle of Bacteriorhodopsin. *Biochemistry (Moscow)* **66**, 1269–1276 (2001).

- 
- [32] Beja, O., J. Spudich, E.N. and Spudich, M. Leclerc und E. DeLong. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
- [33] J.C.Venter und et.al. Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- [34] Spudich, J., D. McCain, K. Nakanishi, M. Okabe, N. Shimizu, H. Rodman, B. Honig und R. Bogomolni. Chromophore/protein interaction in bacterial sensory rhodopsin and bacteriorhodopsin. *Biophys. J.* **49**, 479–483 (1986).
- [35] Smith, S. O., A. Myers, R. Mathies, J. Pardo, C. Winkel, E. van den Berg und J. Lugtenburg. Vibrational analysis of the all-*trans* retinal protonated Schiff base. *Biophys. J.* **47**, 653–664 (1985).
- [36] Braiman, M., T. Mogi, T. Marti, L. Stern, H. Khorana und K. Rothschild. Vibrational Spectroscopy of Bacteriorhodopsin Mutants: Light-Driven Proton Transport Involves Protonation Changes of Aspartic Acid Residues 85, 96, and 212. *Biochemistry* **27**, 8516–8520 (1988).
- [37] Smith, S., J. Lugtenburg und R. Mathies. Determination of retinal chromophore structure in bacteriorhodopsin with resonance Raman spectroscopy. *Journal of Membrane Biology* **85**, 95–109 (1985).
- [38] Großjean, M. F. und P. Tavan. Wavelength regulation in bacteriorhodopsin and halorhodopsin: a PPP-MRD-CI study of retinal dyes. *J. Chem. Phys.* **88**, 4884–4896 (1988).
- [39] van den Berg, R., Du-Jeon-Jang, H. C. Bitting und M. A. El-Sayed. Subpicosecond resonance Raman spectra of the early intermediates in the photocycle of bacteriorhodopsin. *Biophys. J.* **58**, 135–141 (1990).
- [40] Schulten, K. und P. Tavan. A mechanism for the light-driven proton pump of Halobacterium halobium. *Nature* **272**, 85–86 (1978).
- [41] Hayashi, S., E. Tajkhorshid und K. Schulten. Structural Changes during the Formation of Early Intermediates in the Bacteriorhodopsin Photocycle. *Biophys. J.* **83**, 1281–1297 (2002).
- [42] Fodor, S. P. A., W. T. Pollard, R. Gebhard, E. M. M. van den Berg, J. Lugtenburg und R. A. Mathies. Bacteriorhodopsin's L<sub>550</sub> intermediate contains a C<sub>14</sub>-C<sub>15</sub> s-*trans*-retinal chromophore. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2156–2160 (1988).
- [43] Fahmy, K., O. Weidlich, M. Engelhard, J. Tittor, D. Oesterhelt und F. Siebert. Identification of the proton acceptor of Schiff base deprotonation in bacteriorhodopsin: A fourier-transform-infrared study of the mutant Asp85 -Glu in its natural lipid environment. *Photochemistry and Photobiology* **56**, 1073–1083 (1992).
- [44] Gerwert, K. und F. Siebert. Evidence for light-induced 13-*cis*, 14-s-*cis* isomerization in bacteriorhodopsin obtained by FTIR difference spectroscopy using isotopically labelled retinals. *EMBO J.* **5**, 805–811 (1986).
- [45] Großjean, M. F., P. Tavan und K. Schulten. Can normal mode analysis reveal the geometry of the L<sub>550</sub> chromophore of bacteriorhodopsin? *Eur. Biophys. J.* **16**, 341–349 (1989).

- [46] Rödiger, C., I. Chizhov, O. Weidlich, und F. Siebert. Time-resolved step-scan Fourier transform infrared spectroscopy reveals differences between early and late M intermediates of bacteriorhodopsin. *Biophys. J.* **76**, 2687–2701 (1999).
- [47] Grzesiek, S. und N. Dencher. Time-course and stoichiometry of light-induced proton release and uptake during the photocycle of bacteriorhodopsin. *FEBS letters* **208**, 337–342 (1986).
- [48] Hampp, N., C. Bräuchle und D. Oesterhelt. Bacteriorhodopsin wildtype and variant aspartate-96  $\rightarrow$  asparagine as reversible holographic media. *Biophys. J.* **58**, 83–93 (1990).
- [49] Ormos, P. Infrared spectroscopic demonstration of a conformational change in bacteriorhodopsin involved in proton pumping. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 473–477 (1991).
- [50] Bashford, D. und K. Gerwert. Electrostatic Calculations of the  $pK_a$ 's of Ionizable Groups in Bacteriorhodopsin. *J. Mol. Biol.* **224**, 473–486 (1992).
- [51] Spassov, V. Z., H. Luecke, D. Bashford und K. Gerwert.  $pK_a$  Calculations Suggest Storage of an Excess Proton in a Hydrogen-Bonded Water Network in Bacteriorhodopsin. *J. Mol. Biol.* **312**, 203–219 (2001).
- [52] Garczarek, F. und K. Gerwert. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature* **439**, 109–112 (2006).
- [53] Smith, S., J. Pardo, P. Mulder, B. Curry, J. Lugtenburg und R. Mathies. Chromophore structure in bacteriorhodopsin's O640 photointermediate. *Biochemistry* **22**, 6141–6148 (1983).
- [54] Tavan, P., K. Schulten und D. Oesterhelt. The effect of protonation and electrical interactions on the stereochemistry of retinal Schiff bases. *Biophys. J.* **47**, 415–430 (1985).
- [55] Zscherp, C. und J. Heberle. Infrared difference spectra of the intermediates L, M, N, and O of the bacteriorhodopsin photoreaction obtained by time-resolved attenuated total reflection spectroscopy. *J. Phys. Chem. B* **101**, 10542–10547 (1997).
- [56] Eichinger, M., P. Tavan, J. Hutter und M. Parrinello. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* **110**, 10452–10467 (1999).
- [57] Schmitz, M. und P. Tavan. On the art of computing the IR spectra of molecules in condensed phase. In E. Starikov, S. Tanaka und J. Lewis (Herausgeber), *Modern methods for theoretical physical chemistry of biopolymers*, Seiten 159–177. Elsevier, Amsterdam (2006).
- [58] Kandt, C. *Theoretische Untersuchungen integraler photosynthetischer Membranproteine*. Dissertation, Ruhr-Universität Bochum (2003).
- [59] Freddolino, P., F. Liu, S. Park, M. Gruebele und K. Schulten. Microsecond Explicit Solvent Molecular Dynamics Simulations of Protein Folding. *Biophysical Journal* **96**, 590–590 (2009).



- 
- [60] Dong, S., M. Loweneck, T. Schrader, W. Schreier, W. Zinth, L. Moroder und C. Renner. A photocontrolled beta-hairpin peptide. *Chem. Eur. J.* **12**, 1114–1120 (2006).
- [61] Winter, R. und F. Noll. *Methoden der Biophysikalischen Chemie*. Teubner Studienbücher, Stuttgart (1998).
- [62] McPherson, A. Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189**, 1–23 (1990).
- [63] Lanyi, J. K. What is the real crystallographic structure of the L photointermediate of bacteriorhodopsin? *Biochim. Biophys. Acta* **1658**, 14–22 (2004).
- [64] Lanyi, J. K. und B. Schobert. Structural changes in the L photointermediate of bacteriorhodopsin. *J. Mol. Biol.* **365**, 1379–1392 (2007).
- [65] Drenth, J. *Principles of Protein X-Ray Crystallography*. Springer Verlag (1999).
- [66] Garman, E. und T. Schneider. Macromolecular Cryocrystallography. *Journal of Applied Crystallography* **30**, 211–237 (1997).
- [67] Burmeister, W. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica* **56**, 328–341 (2000).
- [68] Royant, A., K. Edman, T. Ursby, E. Pebay-Peyroula, E. Landau und R. Neutze. Helix deformation is coupled to vectorial proton transport in the photocycle of bacteriorhodopsin. *Nature* **406**, 645–648 (2000).
- [69] Lanyi, J. und B. Schobert. Mechanism of Proton Transport in Bacteriorhodopsin from Crystallographic Structures of the K, L, M1, M2, and M2' Intermediates of the Photocycle. *Journal of Molecular Biology* **328**, 439–450 (2003).
- [70] Kouyama, T., T. Nishikawa, T. Tokuhisa und H. Okumura. Crystal Structure of the L Intermediate of Bacteriorhodopsin: Evidence for Vertical Translocation of a Water Molecule during the Proton Pumping Cycle. *Journal of Molecular Biology* **335**, 531–546 (2004).
- [71] Kaptein, R., R. Boelens, R. M. Scheek und W. F. van Gunsteren. Protein Structure from NMR **27**, 5389–5395 (1988).
- [72] van Laarhoven, P. und E. Aarts. *Simulated Annealing: Theory and Applications* (1987).
- [73] Kirkpatrick, S., C. D. Gelatt und M. P. Vecchi. Optimization by Simulated Annealing **220**, 671–680 (1983).
- [74] Patzelt, H., B. Simon, A. terLaak, B. Kessler, R. Kühne, P. Schmieder, D. Oesterheld und H. Oschkinat. The structures of the active center in dark-adapted bacteriorhodopsin by solution-state NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9765–9770 (2002).
- [75] Demtröder, W. *Laserspektroskopie: Grundlagen und Techniken*. Springer (2007).
- [76] Seel, M., E. Wildermuth und W. Zinth. A multichannel detection system for application in ultra-fast spectroscopy. *MEASUREMENT SCIENCE AND TECHNOLOGY* **8**, 449–452 (1997).

- [77] Spörlein, S., H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth und J. Wachtveitl. Ultrafast spectroscopy reveals sub-nanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7998–8002 (2002).
- [78] Gremlich, H. und B. Yan. *Infrared and Raman Spectroscopy of Biological Materials*. Marcel Dekker (2001).
- [79] Goormaghtigh, E., V. Cabiaux und J. Ruyschaert. Physicochemical methods in the study of biomembranes. *Subcellular Biochemistry* **23**, 329–403 (1994).
- [80] Gerwert, K. *Transduktion der Lichtenergie in Protonen-Transfer-Reaktionen beim Bacteriorhodopsin: Eine Untersuchung mit Hilfe der zeitaufgelösten IR- und statischen FTIR-Differenzspektroskopie*. Dissertation, Albert-Ludwigs-Universität Freiburg im Breisgau (1985).
- [81] Alshuth, T. *Kinetische und strukturelle Untersuchungen am Chromophor von Bacteriorhodopsin mit Hilfe zeitaufgelöster Resonanz-Raman-Spektroskopie*. Dissertation, Goerg-August Universität Göttingen (1985).
- [82] Tavan, P. Quantum chemical normal mode analysis for interpretation of vibrational spectra of biomolecules: state of the art. In J. C. Merlin, S. Turrell und J. P. Huvenne (Herausgeber), *Spectroscopy of Biological Molecules, 6th European Conference on the Spectroscopy of Biological Molecules, 3–8 September, 1995, Villeneuve d'Ascq, France*, Seiten 3–6. Kluwer Academic Publishers, Dordrecht (1995).
- [83] Hohenberg, P. und W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B870 (1964).
- [84] Kohn, W. und L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
- [85] Frisch, M. J., G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson *et al.* *Gaussian 92, Revision F.2*. Gaussian, Inc., Pittsburgh PA (1993).
- [86] Mathias, G., M. Eichinger, H. Carstens, M. Stork, A. Weiss *et al.* EGO-MMII users guide. Lehrstuhl für BioMolekulare Optik, Ludwig Maximilian Universität München, Oettingenstrasse 67, D-80538 München. In preparation.
- [87] Kirchner, B., A. Seitsonen und J. Hutter. *CPMD Version 3.7.2*. MPI für Festkörperforschung und IBM Zurich Research Laboratory (2003).
- [88] Nonella, M., G. Mathias, M. Eichinger und P. Tavan. Structures and Vibrational Frequencies of the Quinones in *Rb. sphaeroides* Derived by a Combined Density Functional/Molecular Mechanics Approach. *J. Phys. Chem. B* **107**, 316–322 (2003).
- [89] Nonella, M., G. Mathias und P. Tavan. Infrared Spectrum of *p*-Benzoquinone in Water Obtained from a QM/MM Hybrid Molecular Dynamics Simulation. *J. Phys. Chem. A* **107**, 8638–8647 (2003).
- [90] Klähn, M., G. Mathias, C. Koetting, M. Nonella, J. Schlitter, K. Gerwert und P. Tavan. IR Spectra of Phosphate Ions in Aqueous Solution: Predictions of a DFT/MM Approach Compared with Observations. *J. Phys. Chem. A* **108**, 6186–6194 (2004).

- 
- [91] Schmitz, M. und P. Tavan. Vibrational spectra from atomic fluctuations in dynamics simulations: I. Theory, limitations, and a sample application. *J. Chem. Phys.* **121**, 12233–12246 (2004).
- [92] Schmitz, M. und P. Tavan. Vibrational spectra from atomic fluctuations in dynamics simulations: II. Solvent-induced frequency fluctuations at femtosecond time-resolution. *J. Chem. Phys.* **121**, 12247–12258 (2004).
- [93] Kandt, C., J. Schlitter und K. Gerwert. Dynamics of Water Molecules in the Bacteriorhodopsin Trimer in Explicit Lipid/Water Environment. *Biophys. J.* **86**, 705–717 (2004).
- [94] Jorgensen, W. L., D. S. Maxwell und J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
- [95] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan und M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**, 187–217 (1983).
- [96] Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, J. K. M. Merz *et al.* A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
- [97] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In J. Buchner und T. Kiefhaber (Herausgeber), *Protein Folding Handbook. Part 1*, Seiten 1170–1195. Wiley-VCH, Weinheim (2005).
- [98] Großjean, M. *Korrelationseffekte in Schwingungsspektren und optischen Spektren von Retinalchromophoren*. Dissertation, Technische Universität München (1988).
- [99] Eichinger, M. *Berechnung molekularer Eigenschaften in komplexer Lösungsumgebung: Dichtefunktionaltheorie kombiniert mit einem Molekularmechanik-Kraftfeld*. Dissertation, Ludwig-Maximilians Universität München, Germany (1999).
- [100] Mathias, G. *Anwendung von QM/MM Hybridmethoden zur Berechnung der Schwingungsspektren von biologischen Chromophoren in situ*. Diplomarbeit, Ludwig-Maximilians-Universität, Fakultät für Physik, München, Germany (2000).
- [101] Babytska, G. *Dichtefunktionaltheorie kombiniert mit einem molekularmechanischen Kraftfeld: Die Berechnung der Schwingungsspektren des Retinalchromophores von Bakteriorhodopsin*. Diplomarbeit, Ludwig-Maximilians-Universität, Fakultät für Physik, München, Germany (2004).
- [102] Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140 (1997).
- [103] Sugita, Y. und Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).
- [104] Grudinin, S., G. Büldt, V. Gordeliy und A. Baumgaertner. Water Molecules and Hydrogen-Bonded Networks in Bacteriorhodopsin — Molecular Dynamics Simulations of the Ground State and the M-Intermediate. *Biophys. J.* **88**, 3252–3261 (2005).

- [105] Ponder, J. W. und D. A. Case. Force Fields for Protein Simulations. *Adv. Protein Chem.* **66**, 27–85 (2003).
- [106] Denschlag, R., M. Lingenheil und P. Tavan. Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria. *Chem. Phys. Lett.* **458**, 244–248 (2008).
- [107] VandeVondele, J., M. Krack, F. Mohamed, M. Parrinello, C. T. and J. Hutter. Quicks-  
tep: Fast and accurate density functional calculations using a mixed Gaussian and  
plane waves approach. *Comp. Phys. Comm.* **167**, 103–128 (2005).
- [108] Schultheis, V., R. Reichold, B. Schropp und P. Tavan. A polarizable force field for  
computing the infra-red spectra of the polypeptide backbone. *J. Phys. Chem. B*  
**112**, 12217–12230 (2008).

# Danksagung

Allen, die zum Erfolg dieser Arbeit beigetragen haben, möchte ich an dieser Stelle ganz herzlich danken:

Ganz besonders gilt dies für meinen Betreuer Prof. Paul Tavan, der nicht nur das hochinteressante Thema vorgeschlagen hat, mir enorm viel über BR erzählt hat, sondern mich stets mit seinem persönlichen Einsatz motiviert und tatkräftig unterstützt hat.

Für die gute Zusammenarbeit und die angenehme Arbeitsatmosphäre danke ich ganz herzlich allen ehemaligen und gegenwärtigen Kollegen der Arbeitsgruppe: Robert Denschlag, Martin Lingenheil, Rudolf Reichhold, Martina Stork, Gerald Mathias, Stephan Maciej, Bernhard Schropp, Verena Schultheiß, Thomas Hirschberger, Benjamin Rief und Sebastian Bauer. Mein ausdrücklicher Dank dafür gilt meinem Zimmerkollegen Bernhard Schropp für eine sehr angenehme Arbeitsklima im Büro und seine stete Diskussionsbereitschaft.

Robert Denschlag danke ich auch für die schöne Einführung in die Theorie der Replica Exchange Simulationen.

Thomas Hirschberger und Sebastian Bauer danke ich ganz herzlich für die Einführung in die Univar-Auswertungen.

Ein ganz herzliches Dankeschön gilt für Stephan Maciej für seine schnelle Unterstützung bei der "Rettungsaktion MIME". Seine Leistung dabei weiss ich hoch zu schätzen!

Ein ganz großer Dank gilt auch Dr. Tobias Schrader für die interessante Diskussionen über die Moleküle, ihre IR-Spektren und Spektrenmessungen.

Meinen neuen Biostatistik-Kolleginnen, Sonja Kratzmair und Chistina Rabe, danke ich ganz besonders dafür, dass sie sich den Einblick in die Biophysik getraut haben und meine Dissertation korrektur gelesen haben. Freidemann danke ich ganz herzlich für seine Unterstützung bei der Entwicklung meiner Dissertation, den wunderschönsten Tee und Kekse. Wilhelm danke ich dafür dass, er meinen Erzählungen über meine Doktorarbeit immer gerne zugehört hat und mich reichlich mit Büchern versorgt hat. Christoff danke ich für seine positive Lebenseinstellung, die mir in den schwierigsten Zeiten immer als Leitfaden gedient hat.

Unvorstellbar wäre das Zustandekommen dieser Dissertation auch ohne Hilfe meiner Familie. Ganz besonders danken möchte ich an dieser Stelle meinem Ehemann für seine liebevolle Unterstützung. Meinem Sohn danke ich für seine täglichen Nachfragen über meine Fortschritte und seine Hilfe bei dem Zustandekommen vieler Bilder dieser Arbeit. Unermeßlicher Dank gilt auch meiner Mutter, die mir geholfen hat, meine Familie, die Promotion und zum Schluß auch noch meinen Job unter einen Hut zu kriegen.

# Lebenslauf

Name: Galina Babitzki, geb Litwinenko

Geburtsdatum: 16.07.1977

Geburtsort: Charkow(Ukraine)

Familienstand: verheiratet, 1 Kind geb. 1998

seit 2008 Biostatistikerin, Roche Diagnostics GmbH

2004–2008 Wissenschaftliche Mitarbeiterin, LMU München

2004 Diplom in Physik

2001 – 2004 Studium der Physik, LMU München

1998 – 2001 Studium der Physik, TU Dortmund

1994 – 1997 Studium der Physik, Nationale Universität Charkiw

1992 – 1994 Technisches Liceum, Charkow

